

CAR-TR-858
CS-TR-3799

N00014-95-1-0521
June 1997

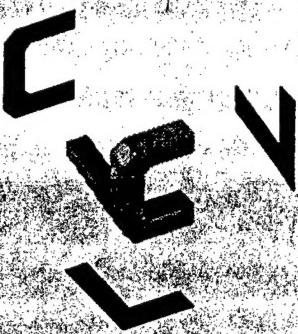
Information-Conserving Object Recognition

Margrit Betke* and Nicholas C. Makris*

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

19980923090

COMPUTER VISION LABORATORY



CENTER FOR AUTOMATION RESEARCH

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742-3275

CAR-TR-858
CS-TR-3799

N00014-95-1-0521
June 1997

Information-Conserving Object Recognition

Margrit Betke[◦] and Nicholas C. Makris*.

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

Abstract

The problem of recognizing objects imaged in complex real-world scenes is examined from a parametric perspective using the theory of statistical estimation. A scalar measure of an object's *complexity*, which is invariant under affine transformation and changes in image noise level, is extracted from the object's Fisher information. The volume of Fisher information is shown to provide an overall statistical measure of the object's *recognizability* in a particular image, while the complexity provides an intrinsically physical measure that characterizes the object in any image. An *information-conserving* method is then developed for recognizing an object imaged in a complex scene. Here the term "information-conserving" means that the method uses all the measured data pertinent to the object's recognizability, attains the theoretical lower bound on estimation error for any unbiased estimate of the parameter vector describing the object, and therefore is statistically optimal. This method is then successfully applied to finding objects imaged in thousands of complex real-world scenes.

[◦] The support of the Office of Naval Research under Contract N00014-95-1-0521 is gratefully acknowledged. Author's new address (starting September 1997): Department of Computer Science, Boston College, Fulton Hall, Chestnut Hill, MA 02167. Email: betke@cs.bc.edu

* N. C. Makris was with the Naval Research Laboratory, Washington, DC 20375. His new address is Department of Ocean Engineering, Massachusetts Institute of Technology, Cambridge, MA 02138.



1 Introduction

Charge-coupled device (CCD) cameras typically produce scene images with extremely low but nonzero noise variance. In fact, for object recognition purposes in computer vision, an initial assumption often is that the noise can be neglected so that the data at each pixel can be regarded as deterministic.

In the present investigation, however, we take an alternative approach that follows a strictly physical interpretation of classical estimation theory. First, we use experimental data to determine the joint probability distribution of the pixel brightness measurements in our CCD images. We use this to construct the likelihood function for any parameter set that is to be estimated given our image data. It is significant that the form of the likelihood function in this physical approach is not arbitrary, but depends upon the probability distribution of the brightness measurements no matter how low the corresponding noise variance is at each pixel, as long as it is nonzero. Moreover, it is the form of this likelihood function, not the level of the noise, that determines the optimal method of recognizing an imaged object.

To emphasize these issues, we show how a scalar measure of an object's *complexity*, which is invariant under affine transformation and changes in image noise level, can be extracted from the determinant of the object's Fisher information matrix. The volume of Fisher information is shown to provide an overall statistical measure of the object's *recognizability* in a particular image, while the complexity provides an intrinsically physical measure that characterizes the object in any image. We then derive a method of recognizing an object in a complex scene that attains the theoretical lower bound on mean-square error for any unbiased estimate of the object's parameter vector, and therefore is by definition *statistically optimal* and *information-conserving*. From a computer vision perspective, we consider the information-conserving property of this estimator to be most significant because it assures that the method uses all the measured data pertinent to the object's recognizability regardless of the noise level. Many popular edge-based methods, for example, discard a significant amount of information pertinent to an object's recognizability and are therefore inherently sub-optimal.

To illustrate our approach, we focus attention in the present paper on the problem of

recognizing objects that are uniquely determined by the six parameters of an affine transformation as well as a seventh parameter that identifies the class of the object. Here, the affine transformation describes rigid body motion and linear distortion of a model object, while the class distinguishes it from other objects with the same affine parameters. For inherently three-dimensional objects, the class must be supplemented by further parameterizations that account for such effects as variation in shading caused by changes in surface orientation with respect to a given source distribution and receiver geometry. For the recognition of flat objects in real world scenes, however, we show that such ancillary parameterizations are unnecessary so long as the object does not have a purely specular surface. This is because the optimal estimator for the affine parameters takes the form of a weighted filter that is invariant to the uniform variations in shading characteristic of such flat objects. This weighting is also necessary to discriminate against image ambiguities that are not explicitly accounted for in classical estimation theory. It is significant that these image ambiguities make the recognition problem inherently nonlinear. A global optimization procedure is therefore necessary to compute the filter output and obtain the optimal estimate.

Our method's performance is evaluated experimentally by applying it to the problem of recognizing traffic signs in images of complicated outdoor scenes. In both our theoretical and experimental analysis, we find that recognizability is strongly dependent upon the object's complexity. We show how this measure becomes analogous to the complexity traditionally referred to in signal processing when the affine transformation is reduced to a 1-D shift in the position of a 1-D object.

2 The statistics of image brightness

Charge-coupled device (CCD) cameras do not output the intensity W of light. Instead, they output a power-transformed intensity on an 8-bit grey-scale which we refer to as image *brightness* $I(x, y)$. The brightness is linearly proportional to $W^{-\gamma}(x, y)$ where γ is a "gamma correction," e.g., $\gamma = 2.2$ [18]. The purpose of this transformation is to correct for the response of cathode-ray tube monitors so that the output of any monitor is proportional to intensity.

Experiments with the CCD video camera used in our vision system indicate that the standard deviation $\sigma(x, y)$ of the output $I(x, y)$ is not only small compared to the mean $m(x, y)$, but, as shown in Figure 1, does not depend on the mean or on position (x, y) . The noise, therefore, is additive and signal-independent, such that $\sigma(x, y) = \sigma$. We speculate that the noise is due to small mechanical vibrations between source and receiver, as well as electronic shot noise. Thermally induced fluctuations of natural light, however, are not a significant cause of errors in our measurements as is shown in Appendix A.

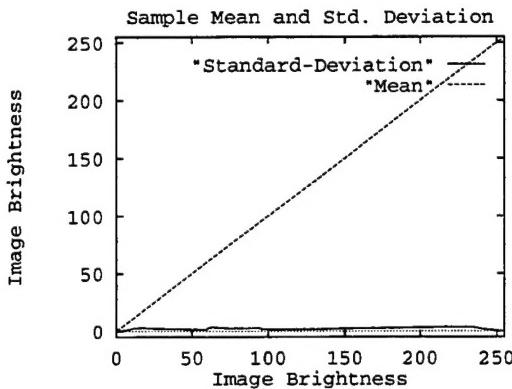


Figure 1: The measured mean and standard deviation of the image brightness I as a function of the mean. The sample standard deviation is signal-independent and obtained by averaging hundreds of images of outdoor scenes. The average standard deviation is 2.65.

Our measured average skew of -0.02 and kurtosis of 2.81 are so close to the corresponding Gaussian values of 0 and 3 , respectively, that our data can be effectively modeled as Gaussian at each pixel. By computation of the sample covariance of brightness between image pixels, our experiments also indicate that the brightness measurements are statistically independent across the pixels.

Let vector \mathbf{I} represent image $I(x, y)$ where the rows of the image are concatenated into one column vector in lexicographic order. Each component I_k of vector \mathbf{I} contains an independent intensity measurement $I(x, y)$ for $1 \leq k \leq MN$. Then the probability density for \mathbf{I} is

$$P(\mathbf{I}) = \frac{1}{(2\pi\sigma^2)^{MN/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{MN} (I_k - m_k)^2\right). \quad (1)$$

3 Recognition as a parameter estimation problem

We use the six-dimensional vector $\mathbf{a} = (x_0, y_0, \theta_0, s_x, s_y, \alpha)$ to describe rigid body motion and linear distortion of an object q in an image with position $\mathbf{x}_0 = (x_0, y_0)$, rotation θ_0 , contractions s_x, s_y , and skew α which vanishes in a rectangular Cartesian coordinate system. For example, suppose the general Cartesian coordinates (x', y') are related to the rectangular Cartesian system (x, y) by the 2-D affine transformation

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}, \quad (2)$$

which can be expressed more succinctly as $\mathbf{x}' = \mathbf{Ax} - \mathbf{x}_0$, where

$$\mathbf{A} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} \cos \theta_0 & \sin \theta_0 \\ -\sin(\theta_0 + \alpha) & \cos(\theta_0 + \alpha) \end{pmatrix}. \quad (3)$$

A model object $q(x', y')$ in some ideal reference frame (x', y') , therefore, appears as a translated, rotated, contracted and skewed object $q(x, y; \mathbf{a})$ in the covariant reference frame (x, y) of an image. The parameters \mathbf{a} are then measured within the image reference frame such that $-\infty < x_0, y_0 < \infty$, $0 \leq \theta_0 \leq 2\pi$, $-\pi/2 \leq \alpha \leq \pi/2$, and $0 < s_x, s_y < \infty$, where dilations occur for $0 < s_x, s_y < 1$ and contractions for $1 < s_x, s_y$.

To account for the possibility that distinct objects may have coincident vectors \mathbf{a} we define an additional parameter ν that identifies the *class* of the object. For example, in traffic sign recognition, a “slow” sign is in a different class from a “yield” sign, although the two may have the same \mathbf{a} .

From the perspective of statistical estimation theory, recognizing an object is the same as estimating the parameters \mathbf{a} and ν .

4 Parameter resolution: Fisher information, recognizability, and the coherence of objects in images

Let us consider the problem of recognizing an object of a given class in some scene. This can equivalently be posed as the problem of estimating the parameter vector \mathbf{a} given the image

data \mathbf{I} . In this case, the likelihood function for \mathbf{a} , given the image data \mathbf{I} , is

$$P(\mathbf{I}|\mathbf{a}) = \frac{1}{(2\pi\sigma^2)^{MN/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{MN} (I_k - m_k(\mathbf{a}))^2\right) \quad (4)$$

where the mean $m_k(\mathbf{a})$ explicitly depends on the parameters to be estimated. The form of this likelihood function, given our CCD data, is very different from that in active radar, laser, and sonar imaging where nonlinear speckle noise is found [13].

The lower bound on the mean-square error in any unbiased estimate $\hat{\mathbf{a}}$ can be expressed as

$$\mathbb{E}[(\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T] \geq \mathbf{J}^{-1}, \quad (5)$$

where the Fisher information matrix \mathbf{J} is defined by

$$J_{ij} = -\mathbb{E}\left[\frac{\partial^2}{\partial a_i \partial a_j} \ln P(\mathbf{I}|\mathbf{a})\right] = \frac{1}{\sigma^2} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left(\frac{\partial m(x, y; \mathbf{a})}{\partial a_i} \frac{\partial m(x, y; \mathbf{a})}{\partial a_j}\right). \quad (6)$$

Here the image mean $m(x, y; \mathbf{a})$ only depends on the parameter vector \mathbf{a} for those pixels $(x, y) \in O^+$ that constitute the expected object $q(x, y; \mathbf{a})$ and any neighboring pixels that are affected by small changes in \mathbf{a} . The Fisher information matrix, therefore, can be reduced to

$$J_{ij} = \frac{1}{\sigma^2} \sum_{(x,y) \in O^+} \frac{\partial q(x, y; \mathbf{a})}{\partial a_i} \frac{\partial q(x, y; \mathbf{a})}{\partial a_j}. \quad (7)$$

It is significant that any of the diagonal entries of the bound can be expressed as

$$\mathbb{E}[(\hat{a}_i - a_i)^2] \geq [\mathbf{J}^{-1}]_{ii} = \frac{\sigma^2}{E} \ell_i^2, \quad (8)$$

where the object *energy*

$$E = \sum_{(x,y) \in O} |q(x, y; \mathbf{a})|^2 \quad (9)$$

and the *coherence scale*

$$\ell_i = \left([\mathbf{J}^{-1}]_{ii} \frac{E}{\sigma^2}\right)^{\frac{1}{2}} \quad (10)$$

for parameter a_i are physical descriptors of the object which are invariant only under rigid body motion. The coherence scale ℓ_i measures the sensitivity of the object to variations in parameter a_i and, therefore, can be interpreted as the width of the object's autocorrelation peak over lags in a_i . An object with relatively high sensitivity to parameter a_i , for

example, will have a relatively narrow autocorrelation peak. The error in estimating parameter a_i , therefore, increases with the corresponding object coherence scale ℓ_i and additive noise variance, but decreases with object energy.

When all parameters are uncoupled and \mathbf{J} is diagonal, the product of n_a coherence scales $\ell_1 \dots \ell_{n_a}$ yields a *coherence volume* that is a scalar measure characterizing the combined n_a -dimensional variations of the object, where n_a is the length of \mathbf{a} . More generally, we define the coherence volume V in terms of the determinant $|\mathbf{J}|$ of the Fisher information matrix by

$$V = \left(\frac{E}{\sigma^2} \right)^{\frac{n_a}{2}} |\mathbf{J}|^{-\frac{1}{2}}. \quad (11)$$

The lower bound can then be written as

$$\mathbf{J}^{-1} = \mathbf{J}_{adj} \left(\frac{\sigma^2}{E} \right)^{n_a} V^2, \quad (12)$$

where \mathbf{J}_{adj} is the adjugate matrix of \mathbf{J} [21].

These coherence scales have compelling physical meanings that will be discussed in the remainder of this section.

From the computer vision perspective, we consider the interpretation of \mathbf{J} as an information measure to be far more useful than its interpretation as the inverse of the theoretical lower bound on estimation error. Our approach and purpose therefore stands apart from Cernuschi-Friis et al.'s [4]. For example, in the type of optical pattern recognition problems encountered with low-variance CCD camera measurements, the associated bounds on object positional resolution fall in the sub-pixel regime, and are somewhat of an overkill. On the other hand, because the volume $|\mathbf{J}|$ of Fisher information is inversely proportional to the limiting mean-square resolutinal volume of the parameters that uniquely specify the object, we consider it to be a scalar measure of the object's *recognizability* in a given image. By Eq. 12 it is seen that there is a direct relationship between this *recognizability* measure and the *physical components* of the Fisher information, namely, the object's *coherence volume* and *energy*. For example, within a given image, where the additive noise variance is uniform, the information volume $|\mathbf{J}|$ only varies with the object's coherence volume and energy. The noise variance, therefore, factors out under variations in object recognizability, regardless the noise level. This shows that it is the physical structure of the likelihood function and not the level of the noise that is most important in properly formulating the recognition problem.

4.1 Position resolution

We first derive the lower bound on the error for any unbiased position estimate of an object with known rotation, contraction and skew. Given the true position $(a_1, a_2) = (x_0, y_0)$, the Fisher information matrix, with elements

$$J_{ij} = \frac{1}{\sigma^2} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left(\frac{\partial q(x - x_0, y - y_0)}{\partial a_i} \frac{\partial q(x - x_0, y - y_0)}{\partial a_j} \right), \quad (13)$$

can be expressed by a spatial “bandwidth matrix” $\mathbf{B} = \sigma^2/E \mathbf{J}$ that characterizes the object. To do so, it is convenient to let the double sum in Eq. 13 be replaced by a continuous double integral so that $q(x, y)$ and $Q(u, v)$ can be defined as Fourier transform pair

$$Q(u, v) = \iint_O q(x, y) e^{-j2\pi(xu+yv)} dx dy \quad (14)$$

and

$$q(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q(u, v) e^{j2\pi(xu+yv)} du dv \quad (15)$$

where $dx dy = (\Delta x)^2$ is the pixel area. The four elements of \mathbf{B} can then be defined by a mean-square bandwidth B_x^2 in x ,

$$B_x^2 = \frac{(2\pi)^2}{(\Delta x)^2 E} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u^2 |Q(u, v)|^2 du dv, \quad (16)$$

a mean-square bandwidth B_y^2 in y ,

$$B_y^2 = \frac{(2\pi)^2}{(\Delta x)^2 E} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v^2 |Q(u, v)|^2 du dv, \quad (17)$$

and a cross-term

$$B_{xy}^2 = B_{yx}^2 = \frac{(2\pi)^2}{(\Delta x)^2 E} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv |Q(u, v)|^2 du dv, \quad (18)$$

with the aid of Parseval’s Theorem

$$(\Delta x)^2 E = \iint_O |q(x, y)|^2 dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |Q(u, v)|^2 du dv. \quad (19)$$

These definitions for the object’s mean-square spatial bandwidth are similar to those introduced for one-dimensional signal waveforms by Gabor [8]. A distinction lies in the positive-semidefinite nature of our object brightness data versus the zero-mean nature of modulated

signal waveform data. As a result, our mean-square bandwidths are defined about zero spatial frequency, as in Ref. [13], while those in the signal processing literature are defined about some average frequency that approximates the carrier frequency for narrowband signals.

Given these definitions and the derivative rule for Fourier transform pairs, the lower bound on position recognition can be expressed as

$$\mathbf{J}^{-1} = \frac{\sigma^2}{E} \mathbf{B}^{-1} = \frac{\sigma^2}{E} \begin{pmatrix} B_y^2 & -B_{xy}^2 \\ -B_{xy}^2 & B_x^2 \end{pmatrix} A_{x_0, y_0}^2, \quad (20)$$

where

$$A_{x_0, y_0} = |\mathbf{B}|^{-\frac{1}{2}} \quad (21)$$

is the *coherence area* of the object, which follows from Eq. 11, where $V = A_{x_0, y_0}$ for this 2-D scenario. For example, the lower bound for estimating x_0 is simply

$$E[(\hat{x}_0 - x_0)^2] \geq J_{x_0}^{-1} = \frac{\sigma^2}{E} \ell_{x_0}^2, \quad (22)$$

where coherence length scale ℓ_{x_0} equals $B_y^2/|\mathbf{B}|$ or $B_y^2 A_{x_0, y_0}^2$, and the lower bound for y_0 is

$$E[(\hat{y}_0 - y_0)^2] \geq J_{y_0}^{-1} = \frac{\sigma^2}{E} \ell_{y_0}^2, \quad (23)$$

where ℓ_{y_0} equals $B_x^2 A_{x_0, y_0}^2$. This analysis provides a 2-D extension of the well-known relationship between a 1-D signal's mean-square bandwidth and the optimal resolution attainable in an estimate of its position [5]. While the coherence length scales ℓ_{x_0} and ℓ_{y_0} could have been obtained directly from Eq. 10 without introducing the mean-square bandwidth concept, this would have circumvented both the historical perspective and an important physical interpretation.

The coherence areas and coherence length scales of two traffic signs, a stop sign and a European no-entry sign, are compared in Figure 2. The stop sign has a much smaller coherence area than the other sign. Its position, therefore, can be resolved much more easily.

The bound on position estimation error is not invariant to changes in object rotation, as is shown in Appendix B by principal component analysis.

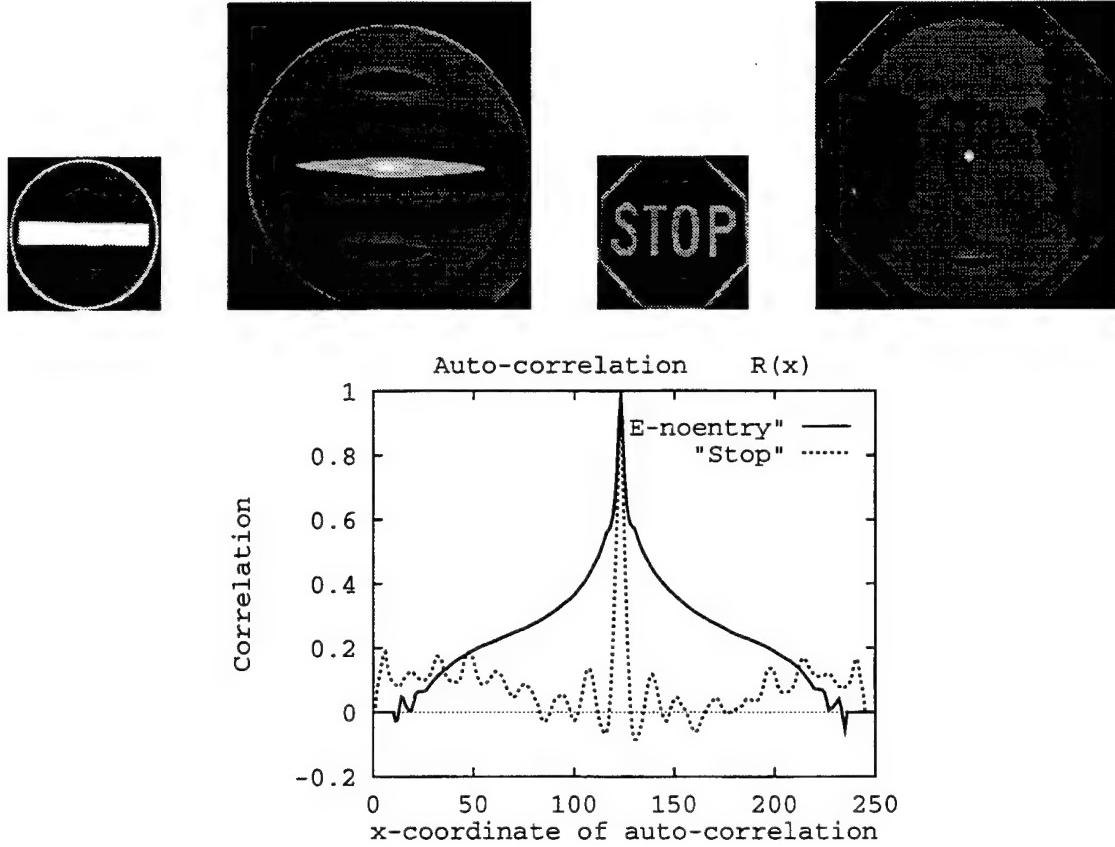


Figure 2: Above, the images of two traffic signs and their 2D-autocorrelation surfaces are shown. The white centers of the autocorrelation surfaces correspond to the coherence areas of the signs. The European no-entry sign's coherence area of 2.2 % of the sign's area is much larger than the stop sign's, which is 0.4 %. This indicates that the position of the stop sign can be resolved more easily than the position of the European no-entry sign. Below are 1D-horizontal slices through the center of the signs' autocorrelation surfaces, where y -positions are fixed and x -positions vary. The stop sign's horizontal position can be resolved better than the European no-entry sign's because of its narrower autocorrelation peak-width and shorter coherence length.

4.2 Angular resolution

Next, assume that only the rotation θ_0 of the object about some point in the image plane is unknown. By Eq. 10, the angular coherence scale for object rotation is

$$\ell_{\theta_0} = \left(\frac{E}{\sum_{(x,y) \in O^+} \left| \frac{\partial g(x,y)}{\partial \theta_0} \right|^2} \right)^{\frac{1}{2}}, \quad (24)$$

This leads to the bound

$$E[(\hat{\theta}_0 - \theta_0)^2] \geq J_{\theta_0}^{-1} = \frac{\sigma^2}{E} \ell_{\theta_0}^2 \quad (25)$$

on angular resolution of the object, which is invariant to changes in object position, since $\frac{\partial x_0}{\partial \theta_0}$ and $\frac{\partial y_0}{\partial \theta_0}$ vanish, but depends on contraction and skew of the object, since E and ℓ_{θ_0} are functions of s_x, s_y , and α .

The angular coherence scales of a stop sign and a European no-entry sign are compared in Figure 3. The European no-entry sign has greater circular symmetry and therefore has a wider angular autocorrelation peak and correspondingly larger angular coherence scale than the stop sign.

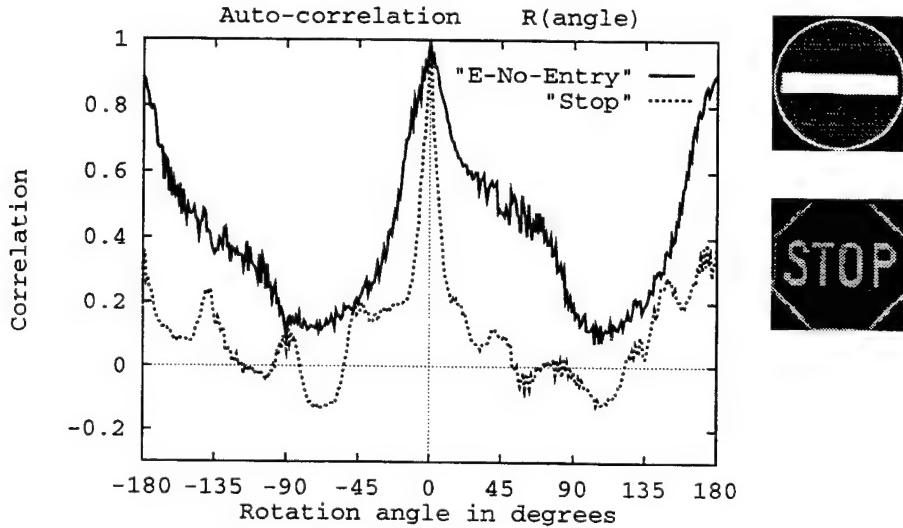


Figure 3: Comparison of angular coherence length scales ℓ_{θ_0} : The European no-entry sign's autocorrelation peak is much wider than the stop sign's, indicating that its rotation is more difficult to resolve.

4.3 Contractional resolution

Finally, assume that only the object's contractional distortions s_x and s_y are unknown. Then, for 2-D parameter vector $(a_1, a_2) = (s_x, s_y)$, where $s_x, s_y > 0$, \mathbf{J} is a 2×2 matrix with elements defined in Eq. 7. The coherence area A_{s_x, s_y} and coherence length scales ℓ_{s_x}, ℓ_{s_y} are then dependent, by Eq. 11, on both diagonal and cross terms of the Fisher information matrix, such that

$$A_{s_x, s_y} = \left(\frac{E}{\sigma^2} \right) |\mathbf{J}|^{-\frac{1}{2}}, \quad (26)$$

$$\ell_{s_x} = \left([\mathbf{J}^{-1}]_{11} \frac{E}{\sigma^2} \right)^{\frac{1}{2}} \quad \text{and} \quad \ell_{s_y} = \left([\mathbf{J}^{-1}]_{22} \frac{E}{\sigma^2} \right)^{\frac{1}{2}}. \quad (27)$$

The bounds for contractional resolution are then

$$E[(\hat{s}_x - s_x)^2] \geq \frac{\sigma^2}{E} \ell_{s_x}^2, \quad (28)$$

and

$$E[(\hat{s}_y - s_y)^2] \geq \frac{\sigma^2}{E} \ell_{s_y}^2. \quad (29)$$

While these bounds are invariant to changes in object position, they are invariant to changes in object rotation only when the contractions s_x and s_y are equal.

The contractional coherence areas and scales of three signs are compared and related to the respective autocorrelation peak widths in Figure 4.

5 The complexity of imaged objects

According to standard usage, an object is considered to be *complex* if it is “composed of elaborately interconnected parts.” We may gather from this that as *complexity* increases so does the number of interconnected parts. These ideas can help us formulate a quantitative definition for the complexity of an imaged object.

Let us first consider two objects of exactly the same dimensions but of different complexities that are imaged in an otherwise empty scene. For example, let the more complex object be a grey-scale Mona Lisa without a picture frame, the less complex object be a blank white canvas of the same dimensions, and the empty background be solid black. Because of their like dimensions, the two objects occupy the same overall area. As may be inferred

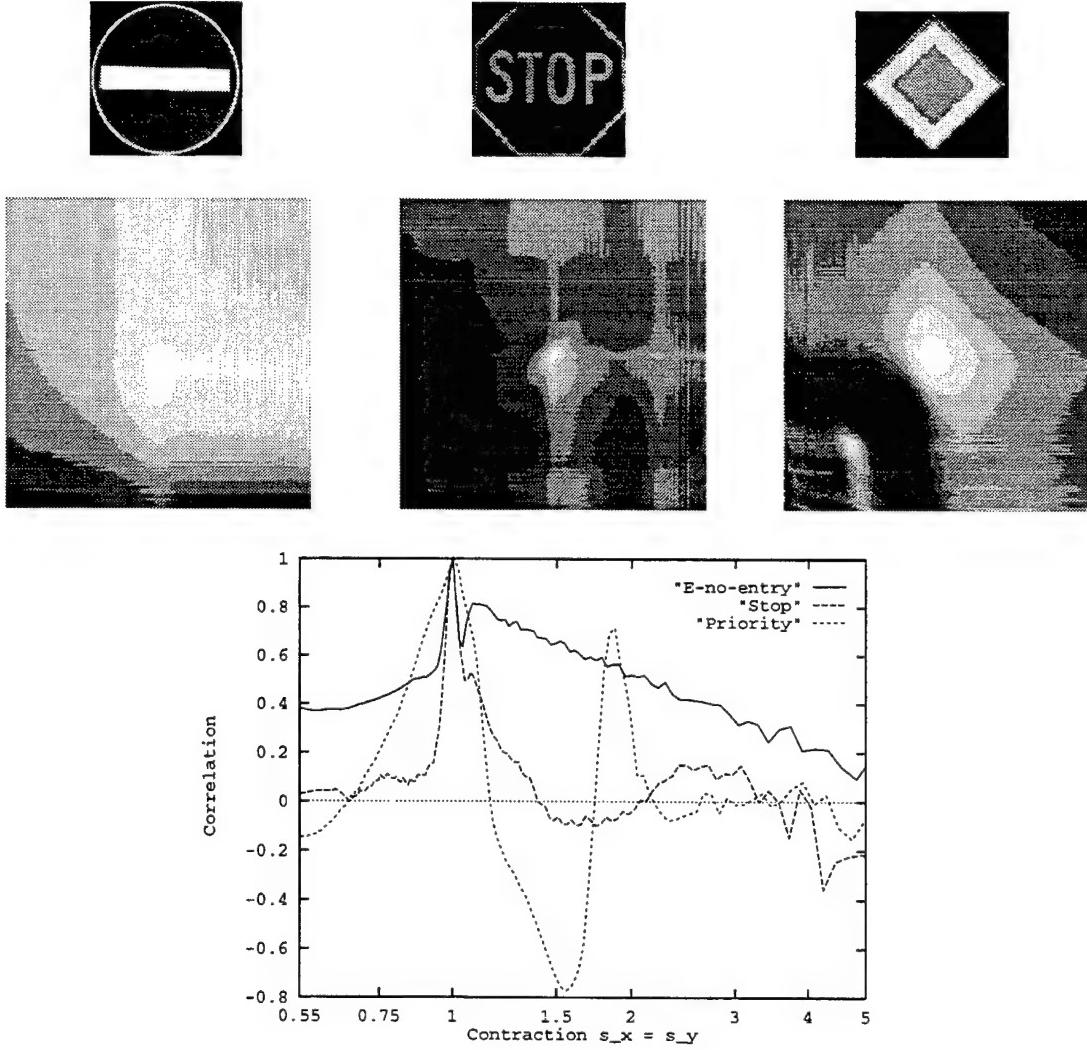


Figure 4: Above, the autocorrelation surfaces of model signs European no-entry, Stop and Priority are shown with contraction parameters s_x and s_y increasing from the lower left to the top right of the surfaces. The white centers of the autocorrelation surfaces are the correlation peaks and correspond to the contractional coherence areas of the signs. The European no-entry sign's contractional coherence area is much greater than the Stop sign's, which means that the contractional parameters s_x and s_y are easier to resolve for the Stop sign. Below, 1-D diagonal slices of the autocorrelation surfaces are shown along the diagonal $s_x = s_y$. Since the peak of the European no-entry sign's autocorrelation is much wider than the stop sign's, the stop sign's size is much easier to resolve than European no-entry sign's.

from their descriptions, however, the two objects have vastly differing coherence areas. Let us regard a coherence area as small if the ratio of it to the overall object area is much less than 1. Then, for example, the Mona Lisa's coherence area will be small, due to its large number "of elaborately interconnected parts," but the number of coherence areas or cells that fit into the Mona Lisa's overall area will be large. Conversely, the coherence area of the blank canvass will not be small, but the number of coherence cells that fit into the blank canvass' overall area will be near unity. We may consider the overall object area as a kind of *outer scale* and the coherence area as a kind of *inner scale* for variations in an object's 2-D position. It is the ratio of such an outer scale to an inner scale that determines the number of coherence cells in the object, also referred to as its degrees of freedom, which can be interpreted as its gain in sensitivity under transformation over the empty object space. By the foregoing argument, this ratio also serves as a quantitative measure of an object's complexity.

Generalizing these concepts, we define the *outer volume* under affine transformation, denoted by S , to be the object area times $2\pi^2$. This is the product of the outer scales for 2-D positional transformation, rotation, 2-D contractions, and skew that are, respectively, the object area A , 2π , unity, and π . The complexity of an object under affine transformation is then the ratio of this outer volume to the coherence volume V defined in Eq. 11, so that

$$C = \frac{S}{V} = A 2\pi^2 \left(\frac{\sigma^2}{E} \right)^{\frac{n_a}{2}} |\mathbf{J}|^{\frac{1}{2}}. \quad (30)$$

The complexities of various traffic signs are compared in Figure 5. As may be expected from a qualitative perspective, signs with inscriptions and human figures have much higher complexities than signs composed only of simple geometric shapes. Our data analysis will later show that the ability to unambiguously resolve such an object increases with the object's complexity.

When the affine transformation is reduced to a 2-D translation, the relevant *positional complexity* becomes

$$C_{x_0, y_0} = \frac{A}{A_{x_0, y_0}}, \quad (31)$$

where the coherence area A_{x_0, y_0} is given in Eq. 21. When the translation is restricted to a single dimension, the above complexity becomes analogous to that used in the signal

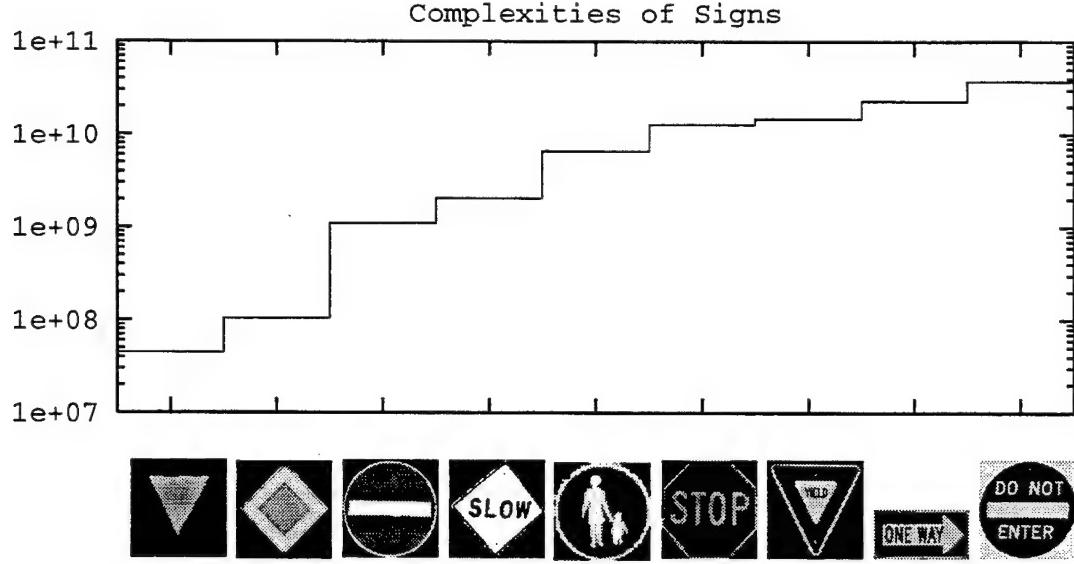


Figure 5: Comparison of complexity C for various traffic signs: Signs with inscriptions and human figures have higher complexity than signs composed only of simple geometric shapes.

processing literature for the analysis of complex waveforms [5, 20].

Similarly, we define the *rotational complexity* of an object by

$$C_{\theta_0} = \frac{2\pi}{\ell_{\theta_0}}, \quad (32)$$

and the *contractional complexity* by

$$C_s = \frac{1}{A_{s_x, s_y}}, \quad (33)$$

where the rotational coherence scale ℓ_{θ_0} is defined in Eq. 24 and the contractional coherence area A_{s_x, s_y} in Eq. 26. These positional, rotational, and contractional complexities of the traffic sign models are plotted in Figure 6 and are consistent with qualitative appraisals of the inherent positional, rotational and contractional symmetries of the signs.

6 Image edges

There is an important connection between the positional Fisher information of an object and “edge-based” recognition. Both require computation of the spatial gradient $(\frac{\partial q(x,y)}{\partial x}, \frac{\partial q(x,y)}{\partial y})$ of the expected object. By Eq. 13, however, the positional Fisher information integrates

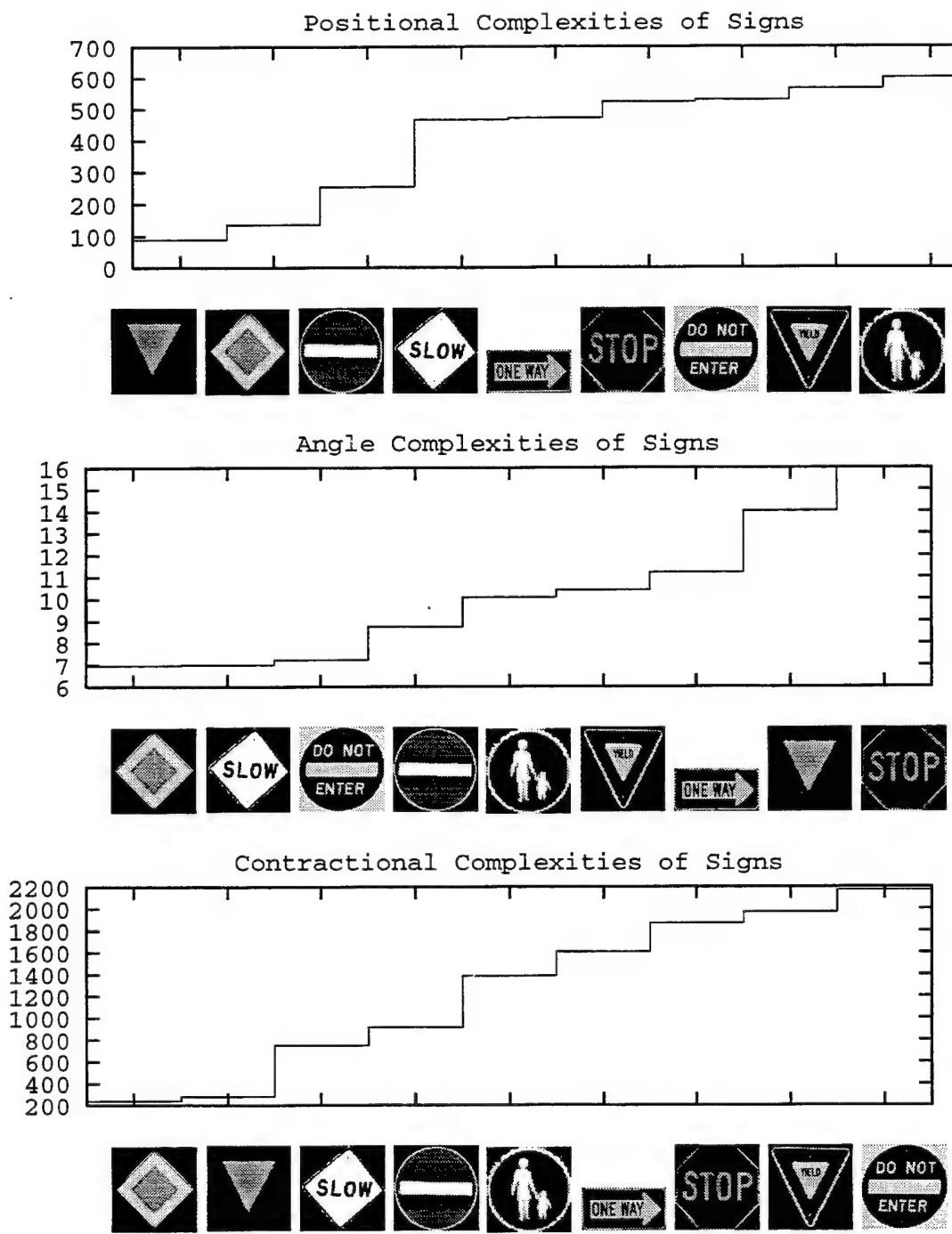


Figure 6: The positional, rotational, and contractual complexities of the traffic sign models.

gradient factors over the entire object. This includes both slowly varying brightness contributions over the entire area of the object as well as rapid variations at edges that comprise a relatively small fraction of the object's overall area. *A priori*, there is no way to judge which of these will make the dominant contribution to the Fisher information. In spite of this basic fact, edge-based recognition methods threshold the gradient magnitude over the object so as to discard all information pertinent to the object's recognizability that is not contained in its edges. The danger in edge-based methods, therefore, is that a potentially larger amount of information may come from slowly varying brightness changes accumulated throughout the object's area than from rapid changes at edges. In this case, edge-based recognition methods are inherently sub-optimal. Conversely, if the predominant positional information about an object is concentrated in its edges, the analysis of Fisher information, coherence scales and complexity remains equally pertinent regardless of the method of recognition. Moreover, the foregoing analysis goes beyond consideration of positional variations, as expressed in terms of the horizontal and vertical gradient components also used in edge methods, but also accounts for the general linear variations permissible in an affine transformation.

Figure 7 illustrates the similarities and differences between a model sign with information concentrated at distinct edges, its derivatives with respect to various affine parameters, and its edge maps.



Figure 7: A stop sign, its partial derivatives with respect to x , y , θ , and $s = s_x = s_y$, and its edge map and thresholded edge map.

7 Maximum likelihood estimation of an object in a scene image

In this section, we derive a method of recognizing an object in a complex scene that attains the theoretical lower bound on mean-square error for any unbiased estimate, and therefore is by definition statistically optimal and information-conserving.

Given the image data \mathbf{I} , and following classical estimation theory, we use the likelihood function of Eq. 4, to derive the maximum likelihood estimate

$$\hat{\mathbf{a}}_{ML} = \arg \max_{\mathbf{a}} P(\mathbf{I}|\mathbf{a}) \quad (34)$$

of the parameters \mathbf{a} . The maximum likelihood estimate $\hat{\mathbf{a}}_{ML}$ can be found by solving the likelihood equation

$$\frac{\partial \ln P(\mathbf{I}|\mathbf{a})}{\partial \mathbf{a}}|_{\mathbf{a}=\hat{\mathbf{a}}_{ML}} = 0. \quad (35)$$

Since

$$\frac{\partial \ln P(\mathbf{I}|\mathbf{a})}{\partial \mathbf{a}} = \frac{\partial}{\partial \mathbf{a}} \left(-\frac{1}{2\sigma^2} \sum_k^{MN} (I_k - m_k(\mathbf{a})) \right)^2, \quad (36)$$

the maximum likelihood estimate $\hat{\mathbf{a}}_{ML}$ is equal to

$$\arg \min_{\mathbf{a}} \frac{1}{2\sigma^2} \left(\sum_{(x,y) \in B} (I(x,y) - m(x,y))^2 + \sum_{(x,y) \in O^+} (I(x,y) - q(x,y; \mathbf{a}))^2 \right), \quad (37)$$

where region B consists of background unrelated to the object while region O^+ is the union of all pixels that contain the expected object $q(x,y; \mathbf{a})$ as well as a slightly perturbed or variational object $q(x,y; \mathbf{a} + \Delta \mathbf{a})$. The first sum in Eq. 37 can be discarded, because the background does not depend on the object properties described by parameter \mathbf{a} . The maximum likelihood estimate is then

$$\hat{\mathbf{a}}_{ML} = \arg \min_{\mathbf{a}} \sum_{(x,y) \in O^+} (I(x,y) - q(x,y; \mathbf{a}))^2. \quad (38)$$

After expanding the square, this reduces to

$$\hat{\mathbf{a}}_{ML} \approx \arg \max_{\mathbf{a}} \sum_{(x,y) \in O^+} I(x,y)q(x,y; \mathbf{a}), \quad (39)$$

because the data energy $\sum_{(x,y) \in O^+} (I(x,y))^2$ is always independent of \mathbf{a} , and, for small perturbations of \mathbf{a} about its true value, the expected object energy $\sum_{(x,y) \in O^+} (q(x,y; \mathbf{a}))^2$ can be taken as a constant independent of \mathbf{a} .

We interpret $q(x,y; \mathbf{a})$, in Eq. 39, as a multidimensional matched filter, which, when evaluated at some particular \mathbf{a}_0 , is referred to as the *replica* $q(x,y; \mathbf{a}_0)$. To find the maximum likelihood estimate $\hat{\mathbf{a}}_{ML}$, therefore, we search for the replica that best matches the object in the image. The value of the parameter vector that corresponds to this best match is the

maximum likelihood estimate. To put it another way, we seek the value of the parameter vector that maximizes the output of the multidimensional matched filter given by the sum in Eq. 39.

To ensure that our filter is not biased by changes in either the data energy or the expected object energy within the local replica window, we employ a local weighting. The output of the resulting weighted multidimensional matched filter

$$r(\mathbf{a}) = \frac{1}{\sigma_I(\mathbf{a})\sigma_q(\mathbf{a})} (A(\mathbf{a}) \sum_{(x,y) \in O} I_q(x,y)q(x,y; \mathbf{a}) - m_I(\mathbf{a})m_q(\mathbf{a})), \quad (40)$$

quantifies how well the measured data in subimage $I_q(x,y)$ matches the replica object in $q(x,y; \mathbf{a})$. Here $A(\mathbf{a})$ is the number of pixels in the replica image $q(x,y; \mathbf{a})$ that have nonzero brightness, and therefore constitute the replica object, while O is the region that contains the replica object, as illustrated in Figure 7. The local variance of subimage $I_q(x,y)$ is

$$\sigma_I^2(\mathbf{a}) = A(\mathbf{a}) \sum_{(x,y) \in O} I_q(x,y)^2 - \left(\sum_{(x,y) \in O} I_q(x,y) \right)^2,$$

the variance of replica image $q(x,y; \mathbf{a})$ is

$$\sigma_q^2(\mathbf{a}) = A(\mathbf{a}) \sum_{(x,y) \in O} q(x,y; \mathbf{a})^2 - \left(\sum_{(x,y) \in O} q(x,y; \mathbf{a}) \right)^2,$$

and the local image means are $m_I(\mathbf{a}) = \sum_{(x,y) \in O} I_q(x,y)$ and $m_q(\mathbf{a}) = \sum_{(x,y) \in O} q(x,y; \mathbf{a})$. It is noteworthy that the weighted multidimensional matched filter is dimensionless, with $|r(\mathbf{a})| \leq 1$ by the Cauchy-Schwartz inequality, so that scene object I_q and replica object q are perfectly correlated when $r(\mathbf{a}) = 1$.

When the estimate $\hat{\mathbf{a}}$ is very close to its true value, small changes in \mathbf{a} lead to negligible changes in m_I , m_q , σ_I , and σ_q , so that these sample means and standard deviations may be taken as locally constant. In this case, the weighted matched filter of Eq. 40 becomes a linear function of the matched filter, the sum in Eq. 39, as demonstrated experimentally in Figure 9, so that the value of \mathbf{a} that maximizes $r(\mathbf{a})$ is the maximum likelihood estimate, where

$$\hat{\mathbf{a}}_{ML} = \arg \max_{\mathbf{a}} r(\mathbf{a}). \quad (41)$$

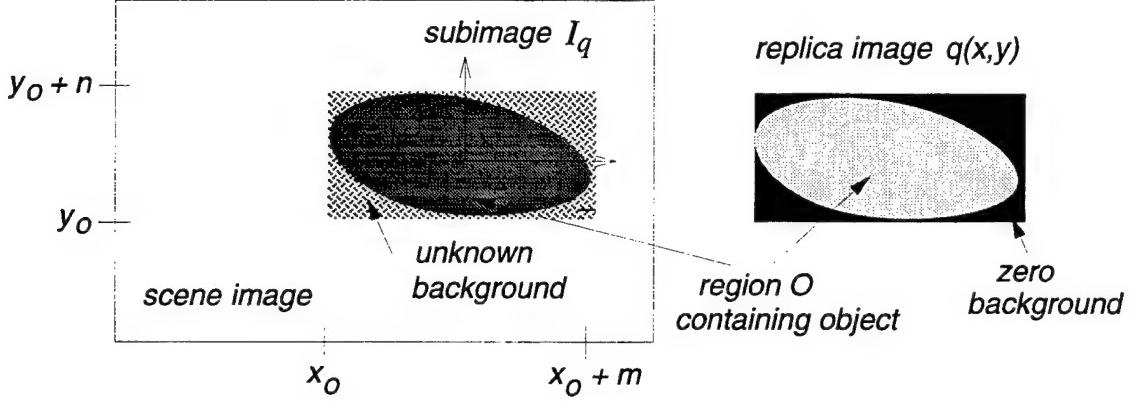


Figure 8: Scene image $I(x,y)$ with subimage $I_q(x,y)$ and replica image $q(x,y)$. Since the replica object may not be exactly rectangular, the portion $I_q(x,y)$ of the scene image that does not overlap the object replica must be removed from the match. To do so, the $m \times n$ replica image $q(x,y)$ is set to zero for pixels not belonging to the replica object. The computation time for any value of r is proportional to the number of nonzero pixels A in the object, which is usually much smaller than the number of pixels in I .

Therefore, while there may be local optima in the weighted matched filter output, the location of its global optimum in the parameter search space corresponds to the maximum likelihood estimate. The maximum likelihood estimate, and hence the weighted matched filter, asymptotically attains the lower bound derived in Eq. 12, and is therefore optimal and information-preserving when the *signal-to-noise-ratio* (SNR) E/σ^2 is high, as it is for the present recognition problem with CCD data. An analytic proof of this can be found in Ref. [14].

However, a more practical experimental proof of our method's statistical optimality is readily provided by inspection of Figure 9. Over the entire global peak, the weighted matched filter, a comparison of noiseless object replicas versus noisy image data, is indistinguishable from the autocorrelation of the coincident noiseless object replicas. While such a result is not surprising, nor does it make the corresponding sub-pixel positional error bounds more relevant, it nonetheless confirms that our approach is information-conserving, and therefore takes full advantage of all the image data pertinent to the object's recognizability.

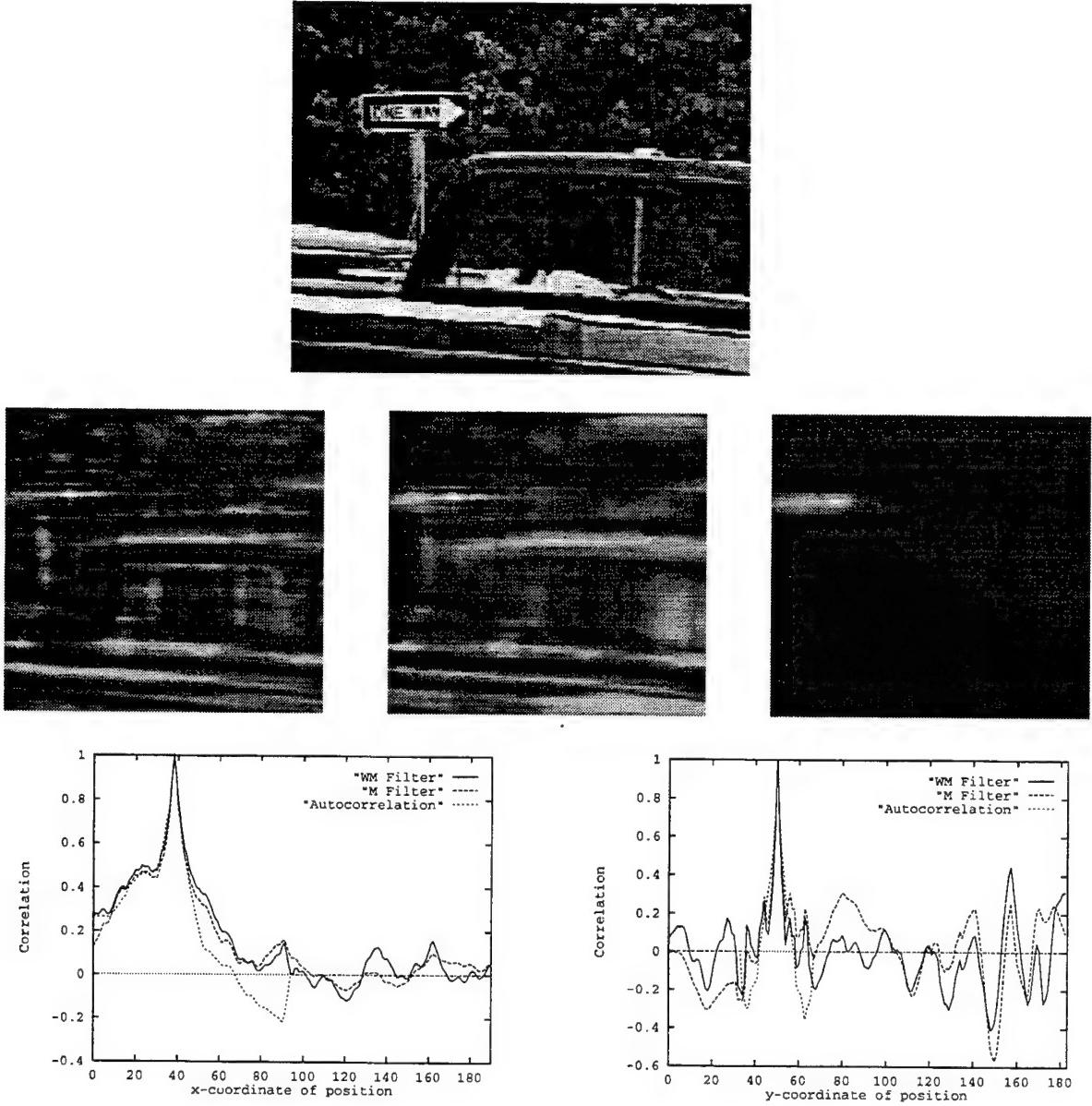


Figure 9: Above, a scene image with a “one way” sign. In the middle, three ambiguity surfaces computed for all possible translations of the “one way” sign replica with fixed angle and scaling parameters. The left surface is computed using the matched (M) filter (Eq. 39), the middle surface is computed using the weighted matched (WM) filter (Eq. 40), and the right surface is the autocorrelation. The correlation peak of the surfaces is a white spot located in the upper left of each plot. Below, horizontal and vertical slices through the global peaks of the ambiguity surfaces. The left graph shows slices along the x -axis of the ambiguity surfaces with the y -coordinate fixed, and the right graph shows slices along the y -axis with the x -coordinate fixed. The methods converge at the true solution.

8 Brightness invariance of flat surfaces

The brightness of an object depends on its reflectance properties, its shape, and its illumination. In particular, the *scene radiance* L of a surface patch centered at world point (X, Y, Z) is proportional to the *image irradiance* or *intensity* W measured at the corresponding pixel (x, y) , such that

$$W(x, y) = g \ L(X, Y, Z), \quad (42)$$

where g is a function of parameters of the imaging system [10]. Since the sensitivity of our imaging system is uniform over the whole image, we can assume that g is constant. The imaging scenario is illustrated in Figure 10.

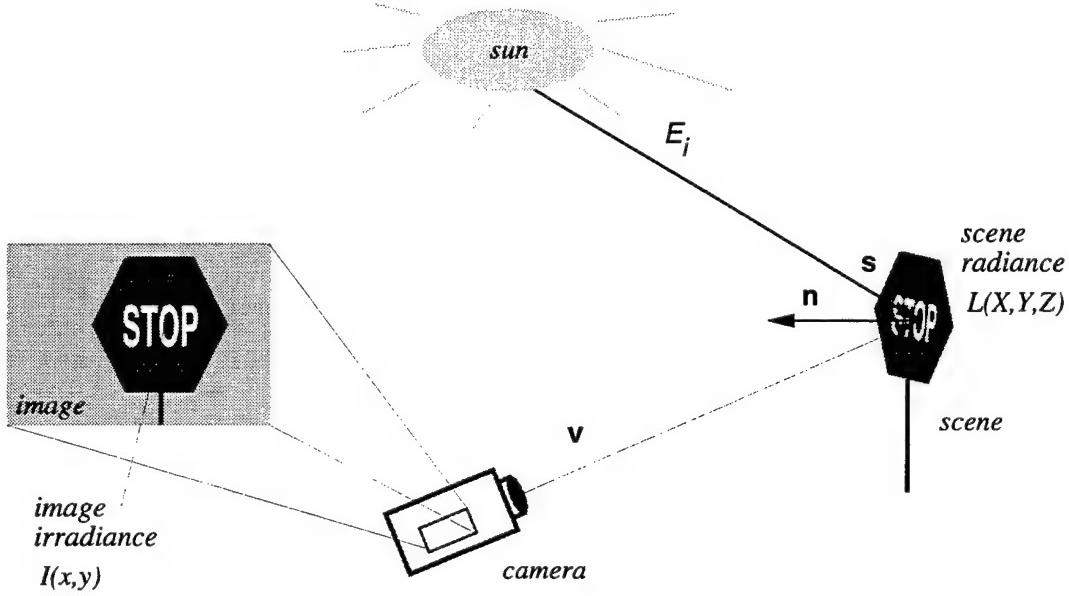


Figure 10: The image irradiance $I(x, y)$ is a function of the corresponding scene radiance $L(X, Y, Z)$, which depends on the source direction s , the viewer direction v , the surface normal n , and the scene irradiance E_i .

The scene radiance is related to the object's bidirectional reflectance distribution function (BRDF) f_r and the source irradiance E_i by

$$L_r(X, Y, Z) = f_r(\mathbf{s}(X, Y, Z), \mathbf{v}(X, Y, Z), X, Y, Z) E_i(\mathbf{s}(X, Y, Z)), \quad (43)$$

where $\mathbf{s}(X, Y, Z)$ is the direction of a collimated light source, and $\mathbf{v}(X, Y, Z)$ is the direction of the camera. For a flat surface, however, the direction of the collimated source is constant over the object such that $\mathbf{s} = \mathbf{s}(X, Y, Z)$. Under the benign assumption that the object's reflectance has directional properties that are separable from its spatial properties, we have

$$f_r(\mathbf{s}, \mathbf{v}(X, Y, Z), X, Y, Z) = f_{r1}(\mathbf{s}, \mathbf{v}(X, Y, Z)) \varrho(X, Y, Z). \quad (44)$$

A special case of this is a Lambertian surface where $f_{r1}(\mathbf{s}, \mathbf{v}) = 1/\pi$ and $\varrho(X, Y, Z)$ is the albedo. If the camera is at least a few object lengths away then its directional variations over the object will be so small that the camera's direction can be considered constant such that $\mathbf{v} = \mathbf{v}(X, Y, Z)$. Then the image brightness $I = W^{-\gamma}$ becomes

$$I = c_r \varrho^{-\gamma}(X, Y, Z), \quad (45)$$

which, to within the constant factor

$$c_r = (g f_{r1}(\mathbf{s}, \mathbf{v}) E_i(\mathbf{s}))^{-\gamma}, \quad (46)$$

is invariant to changes in the geometry of the source, receiver and object. It is noteworthy that in the case of a Lambertian surface, the above result is valid regardless of whether $\mathbf{v}(X, Y, Z)$ is effectively constant or not. This is significant because many real-world surfaces exhibit Lambertian behavior, and traffic signs in particular are designed to have Lambertian reflectance properties.

By distributivity, these results are easily extended to a hemispherical distribution of distant sources, such as the sky, so that the image brightness of the flat object remains invariant to changes in the geometry of the source, receiver and object to within the constant factor c_r .

9 Recognition of flat objects

The output of the weighted matched filter, given in Eq. 40, is invariant to linear transformations of image brightness of the form

$$I'(x, y) = c_1 I(x, y) + c_2, \quad (47)$$

where c_1 and c_2 are scalar constants as shown in Appendix C. But the analysis of the previous section shows that, to within a scalar factor, the image brightness of a flat object remains invariant to changes in scene shading brought upon by changes in the geometry of the source, receiver and object. The output of our weighted matched filter, therefore, is invariant to such changes in scene shading, as is our optimal estimate of the parameters \mathbf{a} , and as is necessary for object recognition.

Uncooperative conditions such as strong shadows and occlusion can change the image irradiance nonuniformly and will cause problems with recognition if they are not accounted for. However, this is not an exclusive weakness of our present formulation, since any approach, for example, systems based on contour or edge detection, will have difficulties in such unpredictable and adverse situations.

10 The traffic sign recognition system

Our method's performance has been evaluated experimentally by applying it to the problem of recognizing traffic signs. This application is very valuable for intelligent vehicles, which can use the recognition results to adjust their speeds or localize themselves in their environments [2]. There are several conference papers on traffic sign recognition, for example, Refs. [1, 6, 11, 17, 19, 25]. Our first results were published in Ref. [3]. Our method stands apart from previous approaches, because it is not restricted to edge detection as Refs. [1, 6, 17], and it does not rely on color information as Refs. [11, 19, 25]. In principle, our approach could be extended by parameterizing color information.

An overview of our traffic sign recognition system is shown in Figure 11. The system has a library of replica models, one for each traffic sign class, which are input along with a scene image. It outputs a description of the recognized traffic sign in the scene image or concludes that the scene does not contain a traffic sign. The system contains three components: a replica generator discussed in Section 11, a weighted matched filter discussed in Section 7, and a parameter perturbation component discussed in Section 12.

The recognition process starts by choosing an arbitrary model class ν and an initial parameter vector \mathbf{a} randomly. The replica generator uses the initial guess of \mathbf{a} to transform

the model image into replica image $q(x, y; \mathbf{a}, \nu)$. This replica is then used by the weighted matched filter $r(\mathbf{a})$ to quantitatively evaluate the match. If the match is poor, meaning it does not satisfy a predetermined threshold, the parameter vector \mathbf{a} is perturbed using simulated annealing, a standard nonlinear optimization method. With the perturbed parameter vector, a new replica is created and tested.

The process of perturbing and evaluating \mathbf{a} is iterated for a fixed amount of time and then repeated for all sign classes. If the best matching replica image $q(x, y; \mathbf{a}^*, \nu^*)$ among all parameters \mathbf{a} and classes ν correlates highly with the scene image, meaning it surpasses a predetermined threshold, it describes the recognized object and is the output of the system. Otherwise, the system outputs that no traffic sign was found in the scene.

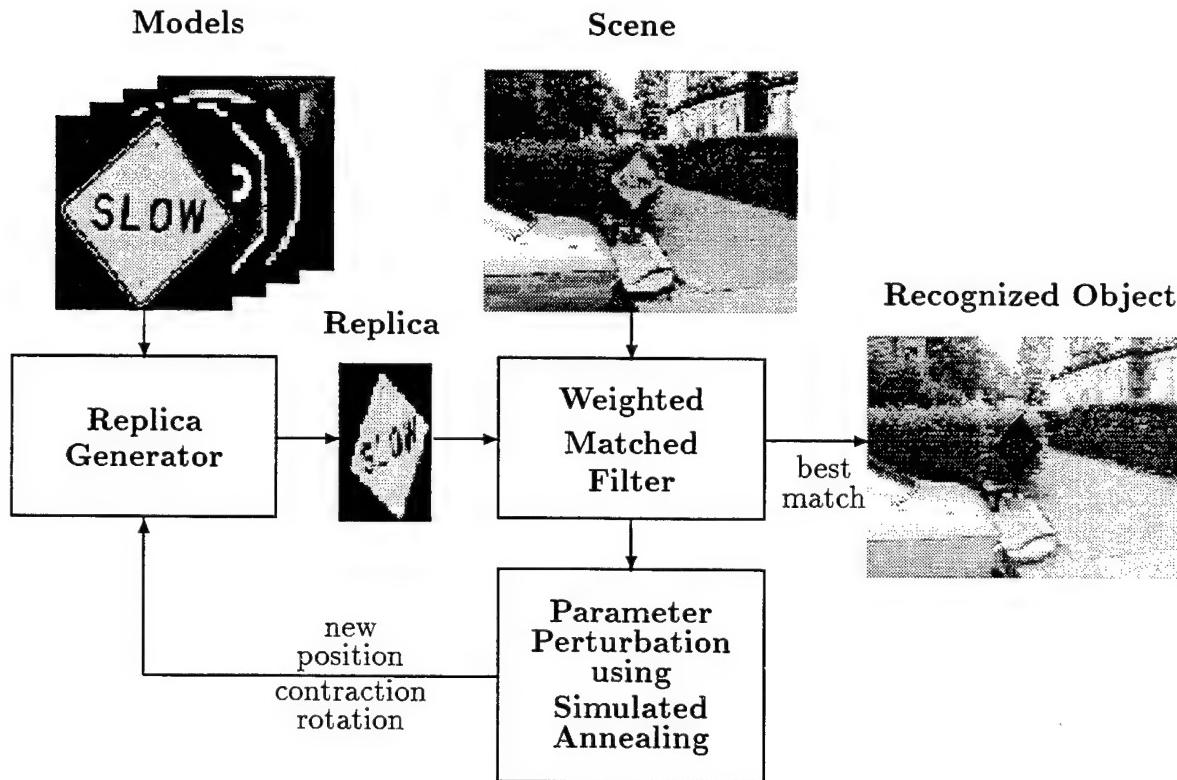


Figure 11: The traffic sign recognition system.

11 Generating replicas from model images

For efficiency reasons, we use five parameters $(x_0, y_0, \theta_0, s_x, s_y)$ to approximate the affine transformation defined in Eq. 3. The skew parameter α is not varied, but set to zero. This

is a valid approximation of transformations that traffic sign images undergo, because the signs are generally fronto-parallel to the image plane or tilted by not more than 45° and are far away from the camera compared to their sizes.

Figure 12 shows how a replica image is generated from a model by subsampling. In this example, the parameters are chosen so that the replica consists only of the circled pixels of the model and additional zero-brightness pixels. For other parameter choices, a four-point interpolation may be necessary to compute the brightness for the replica. Examples of other replicas are shown in Figure 13. Our method computes the replica very quickly by sweeping over the model image only once. The time for creating a replica image from a $n \times m$ model is $O(nm)$.

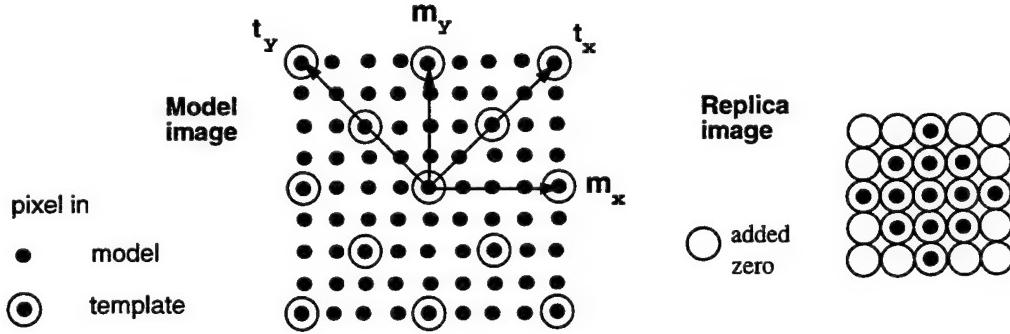


Figure 12: A 5×5 replica image is obtained from a 9×9 model image using contraction parameters $s_x = s_y = 2$ and rotation parameter $\theta_0 = 45^\circ$. For the transformation, the pixels along diagonal vectors t_x and t_y are rotated by 45° and become aligned with the coordinate system of the replica. Similarly, the pixels along vectors m_x and m_y are rotated by 45° and become aligned with the diagonals of the replica. Pixels of zero brightness are added in the replica image where necessary.

12 The simulated annealing algorithm

Since the space of possible solutions of the recognition problem is extremely large, an exhaustive search of this space is computationally too expensive. We therefore base our recognition method on simulated annealing, a popular search technique for solving nonlinear optimization problems [15, 12], which has been applied to many computer vision problems, e.g., [7]. Its name originates from the process of slowly cooling molecules to form a perfect crystal. The analogue to this cooling process is an iterative search process, controlled by a decreasing

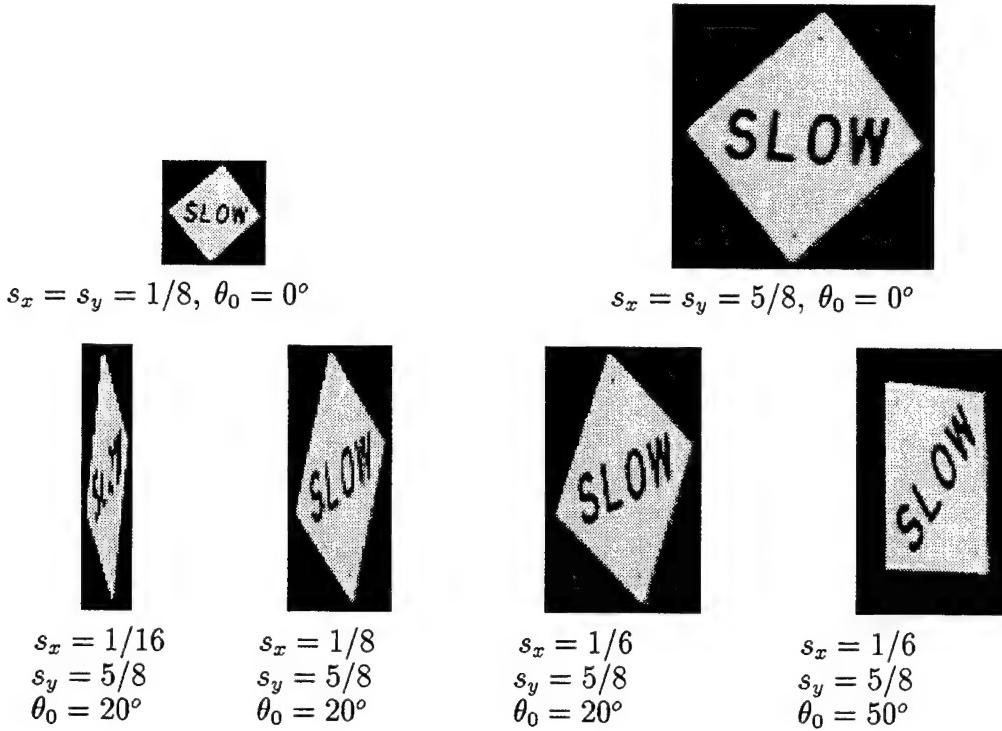


Figure 13: Six replica images of a “slow” sign, obtained by sampling the “slow” sign model at various sampling rates and degrees of rotation.

“temperature” parameter. At each iteration j , the algorithm generates a replica $q(x, y; \mathbf{a}, \nu)$ as described in Section 11. A new test value $a_{\text{test}}^{(j)}$ for parameter a at iteration j is created by

$$a_{\text{test}}^{(j)} = a^{(j-1)} + \Delta a^{(j)}, \quad (48)$$

where $a^{(j-1)}$ is the previous value of a and step $\Delta a^{(j)}$ is a random variable that is uniformly distributed within some interval $[-A, A]$. The step bound A is determined experimentally. To properly deal with image boundaries of a scene image, the j -th test value for the center (x_0, y_0) of a replica of width w_x and height w_y is computed by

$$\begin{aligned} x_{0,\text{test}}^{(j)} &= (x_0^{(j-1)} + \Delta x_0^{(j)} - w_x^{(j-1)}) \bmod (m_I - w_x^{(j-1)}) + w_x^{(j-1)} \\ y_{0,\text{test}}^{(j)} &= (y_0^{(j-1)} + \Delta y_0^{(j)} - w_y^{(j-1)}) \bmod (n_I - w_y^{(j-1)}) + w_y^{(j-1)}. \end{aligned} \quad (49)$$

This definition avoids “attracting” the replica to the rim or corners of the scene image during the search.

At each iteration, the test values for the rotation and contraction parameters are used to create a replica image and correlate it with the scene image at the test location. If

the weighted matched filter output $r_{\text{test}}^{(j)}$ increases over the previous output $r^{(j-1)}$, the test parameter values are accepted since a better match is found, i.e.,

$$\text{if } r_{\text{test}}^{(j)} \geq r^{(j-1)} \quad \text{then } a^{(j)} := a_{\text{test}}^{(j)} \quad (50)$$

for each parameter a . If the current match is worse than the previous match, i.e., $r_{\text{test}}^{(j)} < r^{(j-1)}$, the test values are accepted if

$$\exp\left(-\frac{r^{(j-1)} - r_{\text{test}}^{(j)}}{T^{(j)}}\right) > \xi, \quad (51)$$

where ξ is randomly chosen to be in $[0, 1]$, $T^{(j)}$ is the temperature parameter in the j -th iteration, and the negative exponent corresponds to the Boltzman distribution for thermal equilibrium. For a sufficient temperature, this allows “jumps” out of local maxima. The cooling schedule for the j -th update of the temperature parameter is

$$T^{(j)} = T_0/j \quad \text{for } 1 \leq j \leq L, \quad (52)$$

where T_0 is the initial temperature and L is the number of iterations during the search. Equation 52 describes the fast converging inverse linear cooling schedule [23]. See Ref. [22] for a thorough comparison of annealing algorithms with finite length cooling schedules. Since, after L iterations, the search may not have yielded the optimal solution, a local exhaustive search is conducted around the best solution found. The best result of the local search among all classes describes the recognized sign, as long as it has a filter output that lies above threshold δ . The pseudo code of the recognition algorithm is shown in Figure 14. The behavior of the parameters during a typical run of the algorithm is shown in Figure 15.

13 Experimental results

Our data consists of more than 3280 scene images, a few of which are shown in Figure 16. The main criterion for the selection of the scene images is to obtain a wide variety of traffic sign scenes, originating from both the U.S. and Europe. The signs in the scenes have different sizes and orientations, are illuminated differently, and have various backgrounds. Some traffic signs are aged and bent, some are painted with graffiti. The model images used to represent the traffic sign classes are shown in Figure 19.

RECOGNITION ALGORITHM (Scene image I , set of models \mathcal{M})

1. Initialize recognition threshold δ .
2. Initialize parameter domains.
3. Initialize step bounds $A_{x_0}, A_{y_0}, A_{s_x}, A_{s_y}$, and A_θ .
4. **For** each model class $\nu \in \mathcal{M}$ **do**
5. Initialize position, contraction, and rotation of replica, e.g., at random.
6. Initialize search length L and temperature T_0 .
7. **For** $j = 1$ to L **do**
8. Update temperature $T^{(j)} := T_0/j$.
9. Pick $x_{0,test}^{(j)}$ and $y_{0,test}^{(j)}$ randomly according to Eqs. 48 and 49.
10. Evaluate correlation $r_{test}^{(j)}$ as a function of
 $x_{0,test}^{(j)}, x_{0,test}^{(j)}, s_x^{(j-1)}, s_y^{(j-1)}$ and $\theta_0^{(j-1)}$ according to Eq. 40.
11. Choose ξ uniformly at random within $[0, 1]$.
12. **If** $\exp(-(r^{(j-1)} - r_{test}^{(j)})/T^{(j)}) > \xi$
 $\text{then } x_0^{(j)} := x_{0,test}^{(j)}, y_0^{(j)} := y_{0,test}^{(j)}, r^{(j)} := r_{test}^{(j)}$, update best replica q_ν^*
 $\text{else } x_0^{(j)} := x_0^{(j-1)}, y_0^{(j)} := y_0^{(j-1)}, r^{(j)} := r^{(j-1)}$.
13. Pick $s_{x,test}^{(j)}, s_{y,test}^{(j)}$ and $\theta_{0,test}^{(j)}$ randomly according to Eq. 48.
14. Create new replica with contractions $s_{x,test}^{(j)}, s_{y,test}^{(j)}$ and rotation $\theta_{0,test}^{(j)}$.
15. Evaluate correlation $r_{test}^{(j)}$ as a function of
 $x_0^{(j)}, y_0^{(j)}, s_{x,test}^{(j)}, s_{y,test}^{(j)}$ and $\theta_{0,test}^{(j)}$ according to Eq. 40.
16. Choose ξ uniformly at random within $[0, 1]$.
17. **If** $\exp(-(r^{(j)} - r_{test}^{(j)})/T^{(j)}) > \xi$
 $\text{then } s_x^{(j)} := s_{x,test}^{(j)}, s_y^{(j)} := s_{y,test}^{(j)}, \theta_0^{(j)} := \theta_{0,test}^{(j)}, r^{(j)} := r_{test}^{(j)}$,
 $\text{update best replica } q_\nu^*$.
18. Optimize q_ν^* by small local parameter perturbations
19. Determine replica q^* with highest correlation among all q_ν^* .
20. **If** correlation for q^* < threshold δ
21. **then** output “No traffic sign found in image I .”
22. **else** output “Traffic sign $q(x, y; x_0^*, y_0^*, \theta_0^*, s_x^*, s_y^*; \nu^*)$ found.”

Figure 14: The recognition algorithm.

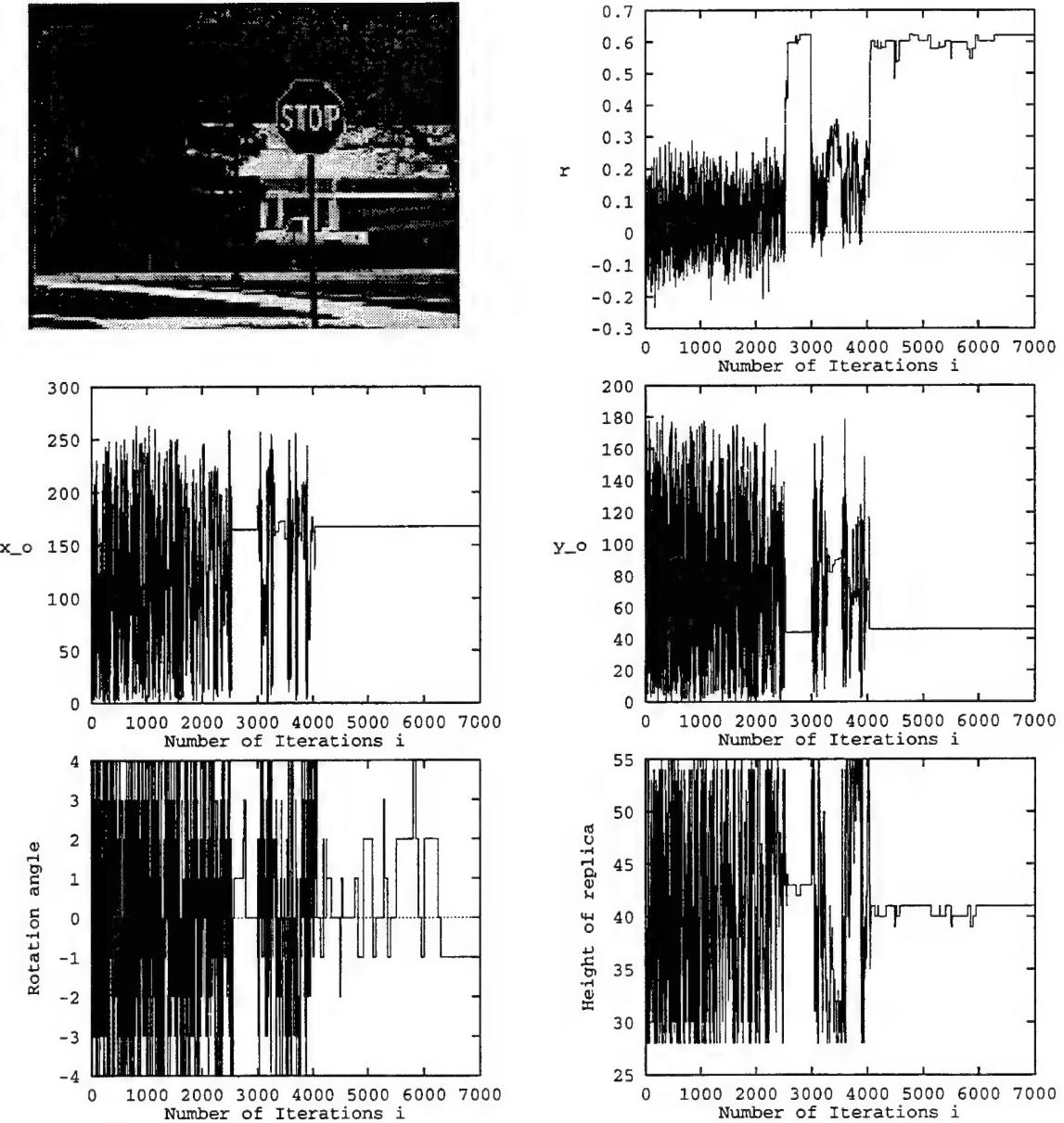


Figure 15: Five graphs illustrating the behavior of the filter output r and the parameters x_0, y_0, θ_0 and s_y during a typical run of the simulated annealing algorithm with a stop sign scene as input. The algorithm is run with the initial parameters reported in Figure 18. The size of the search space is ca. 60 million. The algorithm takes ca. 18 min. on a Sun SPARC station 5. The stop sign is almost recognized at iteration 2802, but the temperature parameter is still too high for convergence. The sign is finally recognized at iteration 4050, after which the parameters are only slightly adjusted.



Figure 16: Some of the images used in the recognition experiments.

The performance of our traffic sign recognition system depends on the complexities of the signs in the images. The system recognizes 94% of the traffic signs correctly and misclassifies 6%, provided that the signs are no smaller than about 8% of the scene's area. These numbers discount mismatches of European signs with their corresponding American signs. For example, replicas of European yield signs do not have any inscriptions, but correlate highly with scenes of American signs with the inscription "yield." Figure 17 shows some recognition results, including scenes with occluded signs and with several traffic signs. (The pseudo-code in Figure 14 is easily modified in line 25 so that several signs in a scene can be found.) The initial parameter values used in the simulated annealing algorithm are listed in Figure 18.

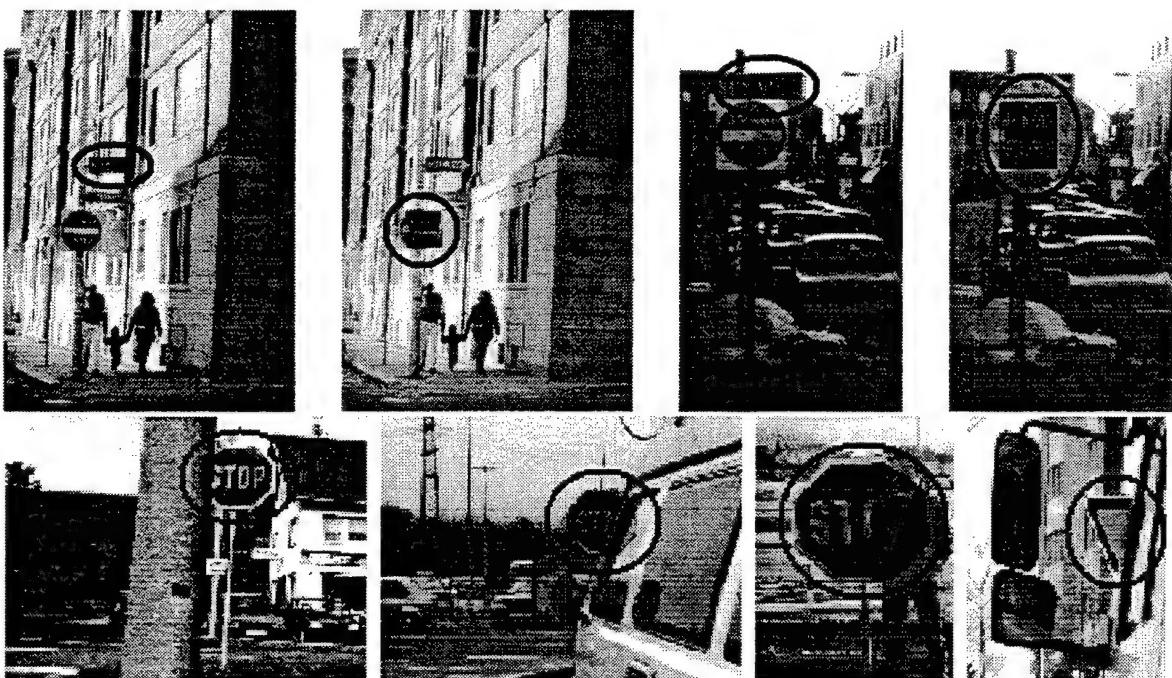


Figure 17: Recognizing multiple and occluded signs: the best matching replica is shown overlying the circled sign in the image.

Recognition threshold:	$\delta = 0.6$
Search length:	$L = 7000$
Initial temperature:	$T_0 = 210$
Initial position:	$(x_0^{(0)}, y_0^{(0)})$ = center of scene
Initial rotation:	$\theta_0^{(0)} = -4^\circ$
Initial contraction:	replica of size 70×70 pixels
Location domain:	$(x_0^{(j)}, y_0^{(j)})$ anywhere in scene
Rotation domain:	$\theta_0^{(j)} \in [-10^\circ, 10^\circ]$
Contraction domain:	$m \times n$ replica with $m, n \in [36 : 140]$ pixels
Step bounds:	$A_{x_0} = A_{y_0} = 50$ pixels, $A_{\theta_0} = 5^\circ$, contraction bounds allow ± 10 pixel steps
Local exhaustive search:	position shift by ± 2 pixels replica contraction by ± 2 pixels replica rotation by $\pm 1^\circ$

Figure 18: Initial parameter values for annealing algorithm.

Figure 19 shows the average best correlation values obtained for each model sign in the experiments. The top graph illustrates the average best correlation for correct matches. For example, replicas created from the footpath sign model match correctly with scene images that contain footpath signs with an average correlation of 0.78. An average of 408 scene images per traffic sign was used, except for the rare “slow” sign for which only a few images could be obtained. Note for comparison that the average correlation for a replica that matches with an arbitrary scene is zero, $E[r] = 0$, while a perfect match yields $r = 1$. The dashed graph in Figure 19 illustrates the average best correlation for scene images that do not contain a traffic sign in the correct class or do not contain a traffic sign at all. For example, the best correlation of a stop sign replica with an arbitrary image that does not contain a stop sign is 0.36 on average. The average is taken over about 2200 scene images per sign class.

False positive matches occur when the best correlating replica is not in the correct class. Figure 20 shows an example of a false positive match where the no-entry sign to be recognized is covered by graffiti and occluded by a nonuniform shadow. Although the sign in the scene can be found, as shown in the left image in Figure 20, the corresponding match yields a filter output that is slightly lower than the output due to the best matching European yield

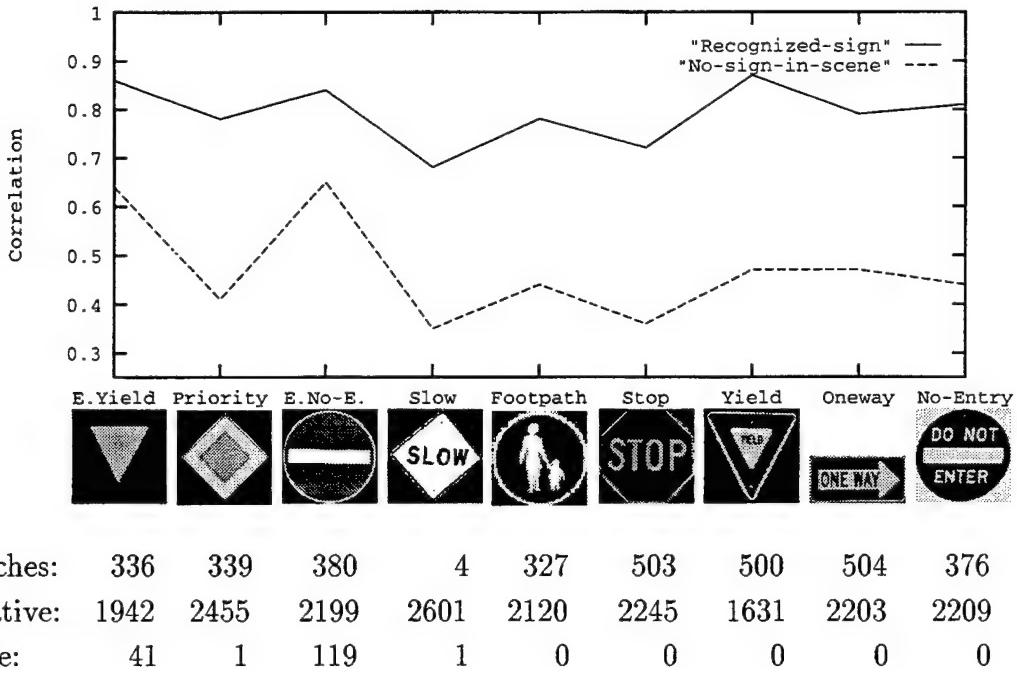
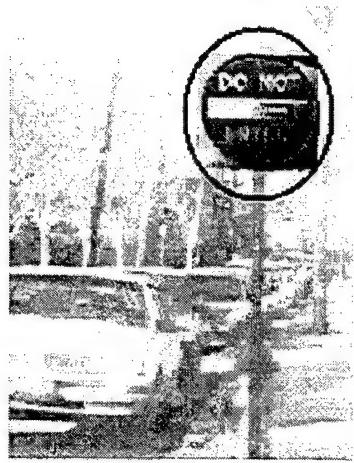
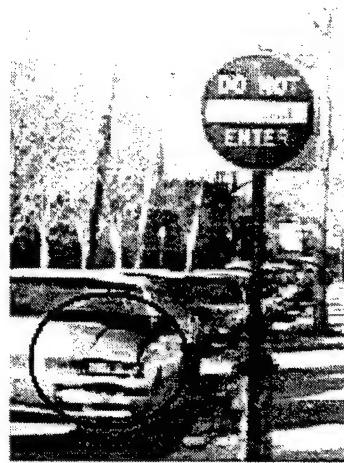


Figure 19: Recognition results for the nine model signs, ordered by their complexities as in Figure 5. The graph plots the average best correlation for scenes with a recognized sign and the average best correlation for scenes without signs. The comparison shows that the correlation for a correct match is high enough to identify the correct class uniquely among the 9 classes, because the false positive correlations are always lower on average.

Underneath the sign models, a table lists the number of correct positive matches and, for scenes without corresponding signs, the number of correct negative and false positive matches. The sum of the entries in each column gives the total number of images used for each model class. False positive matches only occur for *low-complexity* signs.



Correct match, $r = 0.51$



False match, $r = 0.56$

Figure 20: Results for an image with a nonuniform shadow. On the left, the final “no-entry” replica is shown overlying the no-entry sign. On the right, the final “yield” replica is shown overlying an arbitrary part of the image.

sign, as shown in the right image in Figure 20. As can be seen from the data in Figure 19, the European no-entry and European yield models generally result in high filter outputs for arbitrary scenes and are therefore responsible for the vast majority of false matches.

The poor performance of the European no-entry, European yield, and priority signs is due to their low complexities, as defined in Eq. 30. Figure 5 shows that their complexities are orders of magnitude lower than the complexities of signs that do not have such simple geometric shapes.

Traffic signs with inscriptions and complicated shapes are generally more complex and therefore more sensitive to positional, angular, and size variations. Such signs are easier to unambiguously recognize than signs with low complexities. This fact can be used a priori in evaluating the cross-class performance of a recognition system.

While most of the models used in our experiments are complex enough for robust recognition, subsequent downsampling, as is necessary in replica contraction, can be detrimental. Additionally, the complexity of a sign in a scene image can be significantly diminished if the sign is occluded by other objects in the scene. We find that the higher the level of occlusion and the smaller the complexity, the more difficult the recognition of the sign becomes.

14 Summary and conclusions

We have developed a general method for object recognition that is information-conserving; it attains the theoretical lower bound on estimation error for any unbiased estimate regardless of the method of estimation, and is therefore statistically optimal. Our work quantitatively shows how information pertinent to an object's recognizability can be lost by many approaches to object recognition that are commonly used but are inherently sub-optimal from a statistical perspective. Moreover, our theoretical foundation provides a framework for quantitative comparisons between different recognition methods and shows under what special circumstances sub-optimal techniques, such as purely edge-based methods, can become optimal.

We have applied our theoretical results to develop a system that has successfully recognized traffic signs in thousands of complex real-world scenes.

In future work, we will extend our approach to nonplanar 3-D objects, using physical models [10, 24, 16] that describe the imaging process and the object's reflectance properties. This extension of our work will lead to new and interesting descriptors that characterize the intrinsic physical properties of imaged 3-D objects.

References

- [1] S. Azami and M. Aoki. Route guidance sign recognition. In *Proc. Intelligent Vehicles Symp.*, pages 338–343, 1995.
- [2] M. Betke and L. Gurvits. Mobile robot localization using landmarks. *IEEE Trans. Robotics and Automation*, 13:251–263, 1997.
- [3] M. Betke and N. C. Makris. Fast object recognition in noisy images using simulated annealing. In *Proc. ICCV*, pages 523–530, 1995. Also MIT AI Memo 1510.
- [4] B. Cernuschi-Frias, D. B. Cooper, Y.-P. Hung, and P. N. Belhumeur. Toward a model-based Bayesian theory for estimating and recognizing parameterized 3-D objects using two or more images taken from different positions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:540–564, 1989.
- [5] J. V. Difranco and W. L. Rubin. *Radar Detection*. Prentice-Hall, 1968.
- [6] G. Ettinger. Large hierarchical object recognition using libraries of parameterized model sub-parts. In *Proc. CVPR*, pages 32–41, 1988.

- [7] N. S. Friedland and A. Rosenfeld. Lobed object delineation using a multipolar representation. Technical Report 2779, University of Maryland, October 1991.
- [8] D. Gabor. Theory of communication. *J. Inst. Electri. Eng.*, 93:429–457, 1946.
- [9] J. W. Goodman. *Statistical Optics*. Wiley, 1985.
- [10] B. K. P. Horn. *Robot Vision*. MIT Press, 1986.
- [11] N. Kehtarnavaz and A. Ahmad. Traffic sign recognition in noisy outdoor scenes. In *Proc. Intelligent Vehicles Symp.*, pages 460–465, 1995.
- [12] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.
- [13] N. C. Makris. A foundation for logarithmic measures of fluctuating intensity in pattern recognition. *Optics Letters*, 20:2012–2014, 1995.
- [14] N. C. Makris. Optimal pattern recognition in speckle noise, 1997. In preparation.
- [15] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Physics*, C-21:734–738, 1953.
- [16] S. K. Nayar and M. Oren. Visual appearance of matte surfaces. *Science*, 267:1153–1156, 1995.
- [17] G. Piccioli, E. De Micheli, and M. Campani. A robust method for road signed detection and recognition. In *Proc. Europ. Conf. Computer Vision*, volume 1, pages 495–500, 1994.
- [18] C. Poynton. “Gamma” and its disguises: the nonlinear mappings of intensity in perception, CRTs, film and video. *J. SMPTE*, 102:1099–1108, 1993.
- [19] L. Priese, R. Lakmann, and V. Rehrmann. Ideogram identification in a realtime traffic sign recognition system. In *Proc. Intelligent Vehicles Symp.*, pages 310–314, 1995.
- [20] W. Siebert. *Circuits, Signals, and Systems*. MIT Press, 1986.
- [21] G. Strang. *Linear Algebra and its Applications*. Academic Press, 1976.
- [22] P. N. Strenski and S. Kirkpatrick. Analysis of finite length annealing schedules. *Algorithmica*, 6:346–366, 1991.
- [23] H. Szu and R. Hartley. Fast simulated annealing. *Physics Letters A*, 122:157–162, 1987.
- [24] K. Torrance and E. Sparrow. Theory for off-specular reflection from rough surfaces. *J. Opt. Soc. Am. A*, 1:1105–1114, 1967.
- [25] Y. Zheng, W. Ritter, and R. Janssen. An adaptive system for traffic sign recognition. In *Proc. Intelligent Vehicles Symp.*, pages 165–170, 1994.

A Signal-dependent fluctuations of natural light are negligible

This section shows that thermally induced fluctuations of natural light are not a significant cause of errors in our measurements. Natural light fluctuates as a circular complex Gaussian random (CCGR) process [9]. The probability density of an $M \times N$ intensity image \mathbf{W} measured from a CCGR field is the gamma distribution

$$P(\mathbf{W}) = \prod_{k=1}^{MN} \frac{1}{\Gamma(\mu)} \left(\frac{\mu}{\sigma_k} \right)^\mu W_k^{\mu-1} \exp \left(-\mu \frac{W_k}{\sigma_k} \right), \quad (53)$$

where the number of coherence cells μ in the intensity average is defined to be the time-bandwidth product $\mu = T\tau$, where T is the coherence or measurement time, and τ is the bandwidth of the light. The expected value of W_k is σ_k , and the variance of W_k is σ_k^2/μ . Given Eq. 53, the probability density for the “gamma-corrected” brightness \mathbf{I} is

$$P(\mathbf{I}) = \prod_{k=1}^{MN} \frac{1}{\Gamma(\mu)} \left(\frac{\mu}{\sigma_k} \right)^\mu I_k^{\mu-1} \exp \left(-\mu \frac{I_k}{\sigma_k} \right), \quad (54)$$

since $I_k = W_k^{\frac{1}{\gamma}}$. For notational convenience, the subscript k is dropped in the following. The expected value of I is

$$\mathbb{E}[I] = \int_0^\infty W^{\frac{1}{\gamma}} P(W) dW = \left(\frac{\sigma}{\mu} \right)^{\frac{1}{\gamma}} \frac{\Gamma(\frac{1}{\gamma} + \mu)}{\Gamma(\mu)},$$

and the variance of I is

$$\text{var}(I) = \mathbb{E}[I^2] - \mathbb{E}[I]^2 = \left(\frac{\sigma}{\mu} \right)^{\frac{2}{\gamma}} \frac{\Gamma(\mu)\Gamma(\frac{2}{\gamma} + \mu) - (\Gamma(\frac{1}{\gamma} + \mu))^2}{(\Gamma(\mu))^2}.$$

An approximation of the mean of I using Stirling’s Formula that $\Gamma(\mu) = \mu^{\mu-\frac{1}{2}} e^\mu (2\pi)^{\frac{1}{2}}$ yields

$$\mathbb{E}[I] \approx \sigma^{\frac{1}{\gamma}} \frac{(\frac{1}{\gamma} + \mu)^{\frac{1}{\gamma} + \mu - \frac{1}{2}} e^{-(\frac{1}{\gamma} + \mu)}}{\mu^{\frac{1}{\gamma}} \mu^{\mu - \frac{1}{2}} e^\mu} \approx \sigma^{\frac{1}{\gamma}} e^{\frac{\frac{1}{\gamma} + \mu - \frac{1}{2}}{\gamma \mu} - \frac{1}{\gamma}} \approx \sigma^{\frac{1}{\gamma}},$$

which holds for large μ . The variance of I can be approximated by

$$\text{var}(I) \approx \frac{\sigma^{\frac{2}{\gamma}}}{\mu \gamma^2},$$

and is therefore a function of the mean, which reveals that the noise arising from circular complex Gaussian random fluctuations in the received field is *signal-dependent*. This is

important for radar and sonar imaging [13], where, due to signal-dependent fluctuation noise, the variance of high intensity measurements can be larger than the mean of low intensity measurements. For fluctuations of natural light, however, the intensity average of the measurements is large enough to reduce the standard deviation to a negligibly small fraction of its mean, as shown in the following example for green light.

Green light has a bandwidth of $\tau_{\text{green}} = 3 \times 10^8 \frac{\text{m}}{\text{s}} (550\text{nm} - 500\text{nm}) / (550\text{nm} \times 500\text{nm}) = 5.45 \times 10^{13} \text{ Hz}$. With exposure time of $T = 1/100 \text{ Hz}$, the number of coherence cells is $\mu = 5.45 \times 10^{11}$. The ratio of the standard deviation of I to the mean of I is approximately

$$\frac{\text{std}(I)}{\mathbb{E}[I]} \approx \sqrt{\frac{\sigma^2}{\mu\gamma^2}} \frac{1}{\sigma^{\frac{1}{\gamma}}} = \frac{1}{\sqrt{\mu\gamma^2}},$$

which is $O(6 \times 10^{-7})$, a negligibly small ratio compared to that actually measured in Section 2. Therefore, the inherent signal-dependent fluctuations of natural light have a negligible effect on our image data.

B The lower bound on position estimation

This section analyzes how the lower bound on position estimation, as derived in Section 4.1, varies with changes in object rotation. Let \mathbf{R} be a two-dimensional orthonormal matrix. The Fisher information can then be expressed in terms of \mathbf{R} and a diagonal matrix \mathbf{D} that consists of \mathbf{R} 's principal components D_{11} and D_{22} ,

$$\mathbf{J} = \frac{E}{\sigma^2} \mathbf{B} = \frac{E}{\sigma^2} \mathbf{R} \mathbf{D} \mathbf{R}^T = \frac{E}{\sigma^2} \mathbf{R} \begin{pmatrix} D_{11} & 0 \\ 0 & D_{22} \end{pmatrix} \mathbf{R}^T. \quad (55)$$

where

$$D_{11} = \frac{B_x^2 - B_y^2}{2} + \frac{1}{2} \sqrt{4B_{xy}^4 + (B_x^2 - B_y^2)^2} \quad (56)$$

$$D_{22} = \frac{B_x^2 - B_y^2}{2} - \frac{1}{2} \sqrt{4B_{xy}^4 + (B_x^2 - B_y^2)^2}. \quad (57)$$

The angle φ that rotates the x - and y -axes of the image into the principal axes given by the object's bandwidth matrix is defined by

$$\tan(2\varphi) = \frac{2B_{xy}^2}{B_x^2 - B_y^2}. \quad (58)$$

The lower bound on position recognition is therefore

$$\mathbf{J}^{-1} = \frac{\sigma^2}{E} \mathbf{R} \mathbf{D}^{-1} \mathbf{R}^T. \quad (59)$$

The lower bound on recognizing the position coordinate x_0 can then be expressed as

$$E[(\hat{x}_0 - x_0)^2] \geq J_{x_0}^{-1} = \frac{\sigma^2}{E} \frac{D_{11} \sin \varphi + D_{22} \cos \varphi}{D_{11} D_{22}} \quad (60)$$

and on recognizing the coordinate y_0 as

$$E[(\hat{y}_0 - y_0)^2] \geq J_{y_0}^{-1} = \frac{\sigma^2}{E} \frac{D_{11} \cos \varphi + D_{22} \sin \varphi}{D_{11} D_{22}}. \quad (61)$$

If the coordinate axes correspond to the principal components of the bandwidth of the object, i.e., $B_x^2 = D_{11}$ and $B_y^2 = D_{22}$, then the error in the x - and y -coordinate of the position is lower bounded by $\sigma^2/(EB_x^2)$ and $\sigma^2/(EB_y^2)$, respectively. Let the total estimation error be the Euclidean distance $\xi = \sqrt{E[(\hat{x}_0 - x_0)^2]^2 + E[(\hat{y}_0 - y_0)^2]^2}$. The total error is then lower bounded by

$$\xi \geq \frac{\sigma^2}{E} \sqrt{D_{11}^2 + D_{22}^2 + 2D_{11}D_{22} \sin 2\varphi}. \quad (62)$$

The bound is smallest if the principal components of the object's bandwidth matrix are aligned with the coordinate axes, i.e., $\varphi = 0$ or $\varphi = \pi/2$. The bound is largest for $\varphi = \pi/4$, for which $E[(\hat{x}_0 - x_0)^2] = E[(\hat{y}_0 - y_0)^2]$.

C Linear invariance of the weighted matched filter

Given the definition in Eq. 40, the output of the weighted matched filter

$$r(I_q, c_1 q + c_2) = \frac{A \sum I_q(x, y)(c_1 q(x, y) + c_2) - (\sum I_q(x, y))(\sum(c_1 q(x, y) + c_2))}{\sqrt{A \sum I_q(x, y)^2 - (\sum I_q(x, y))^2} \sqrt{A \sum (c_1 q(x, y) + c_2)^2 - (\sum(c_1 q(x, y) + c_2))^2}}$$

describes how well a linearly transformed replica $c_1 q(x, y) + c_2$ matches with the measured data in subimage $I_q(x, y)$. The numerator of $r(I_q, c_1 q + c_2)$ is

$$c_1 \left(A \sum I_q(x, y)q(x, y) - \sum I_q(x, y) \sum q(x, y) \right),$$

and the second square root in the denominator is

$$(A \sum (c_1^2(q(x, y))^2 + 2c_1 c_2 q(x, y) + c_2^2) - c_1^2(\sum q(x, y))^2 - 2c_1 A c_2 \sum q(x, y) - (Ac_2)^2)^{1/2},$$

which yields

$$c_1 \sqrt{A \sum (q(x, y))^2 - (\sum q(x, y))^2}.$$

Since

$$\begin{aligned} r(I_q, c_1 q + c_2) &= \frac{c_1 (A \sum I_q(x, y) q(x, y) - (\sum I_q(x, y)) (\sum q(x, y)))}{\sqrt{A \sum I_q(x, y)^2 - (\sum I_q(x, y))^2} c_1 \sqrt{A \sum (q(x, y))^2 - (\sum q(x, y))^2}} \\ &= r(I_q, q), \end{aligned}$$

the weighted matched filter is invariant to linear transformations of image brightness and Eq. 47 holds.

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)			2. REPORT DATE June 1997		3. REPORT TYPE AND DATES COVERED Technical Report	
4. TITLE AND SUBTITLE Information-Conserving Object Recognition			5. FUNDING NUMBERS N00014-95-1-0521			
6. AUTHOR(S) Margarit Betke and Nicholas C. Makris						
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Center for Automation Research University of Maryland College Park, MD 20742-3275			8. PERFORMING ORGANIZATION REPORT NUMBER CAR-TR-858 CS-TR-3799			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research 800 North Quincy Street, Arlington, VA 22217-5660			10. SPONSORING/MONITORING AGENCY REPORT NUMBER			
 Advanced Research Projects Agency 3701 North Fairfax Drive, Arlington, VA 22203-1714						
11. SUPPLEMENTARY NOTES						
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution unlimited.				12b. DISTRIBUTION CODE		
13. ABSTRACT (Maximum 200 words) The problem of recognizing objects imaged in complex real-world scenes is examined from a parametric perspective using the theory of statistical estimation. A scalar measure of an object's <i>complexity</i> , which is invariant under affine transformation and changes in image noise level, is extracted from the object's Fisher information. The volume of Fisher information is shown to provide an overall statistical measure of the object's <i>recognizability</i> in a particular image, while the complexity provides an intrinsically physical measure that characterizes the object in any image. An <i>information-conserving</i> method is then developed for recognizing an object imaged in a complex scene. Here the term "information-conserving" means that the method uses all the measured data pertinent to the object's recognizability, attains the theoretical lower bound on estimation error for any unbiased estimate of the parameter vector describing the object, and therefore is statistically optimal. This method is then successfully applied to finding objects imaged in thousands of complex real-world scenes.						
14. SUBJECT TERMS Object recognition, statistical estimation, object complexity, traffic signs					15. NUMBER OF PAGES 43	
					16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL			

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to stay *within the lines* to meet optical scanning requirements.

Block 1. Agency Use Only (Leave blank).

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered.

State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract	PR - Project
G - Grant	TA - Task
PE - Program Element	WU - Work Unit
	Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement.

Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."

DOE - See authorities.

NASA - See Handbook NHB 2200.2.

NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.

DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.

NASA - Leave blank.

NTIS - Leave blank.

Block 13. Abstract. Include a brief (*Maximum 200 words*) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (*NTIS only*).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.